DOCUMENT RESUME

ED 440 132 TM 030 744

AUTHOR Aschbacher, Pamela R.

TITLE Developing Indicators of Classroom Practice To Monitor and Support School Reform. CSE Technical Report.

INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.; Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

REPORT NO CSE-TR-513

PUB DATE 1999-12-00

NOTE 117p.

CONTRACT R305B60002

AVAILABLE FROM Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California at Los Angeles, 300 Charles E. Young Dr. North, Los Angeles, CA 90095-1522 ($7.50). Tel: 310-266-1532. For full text: http://www.cse.ucla.edu/CRESST/Reports/TECH513.PDF.

PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143)

EDRS PRICE MF01/PC05 Plus Postage.

DESCRIPTORS *Educational Change; Educational Practices; *Elementary School Teachers; *Middle School Teachers; Middle Schools; Primary Education; *School Restructuring; Scoring Rubrics; Student Evaluation; Urban Schools

IDENTIFIERS Educational Indicators; *Monitoring; Reform Efforts

ABSTRACT

This report describes the development of indicators of classroom practice for monitoring and improving the quality of school reforms. The work entailed development of a rubric to rate key factors of classroom practice based on assignments and samples of student work. This approach was used to describe the intellectual challenge of class assignments, the alignment of tasks with learning goals and grading criteria, clarity of criteria for success, and provision of informative feedback to students. It also compared teacher judgments of student work with external rate judgments using a school district's standards-based rubric. Twenty-four teachers from elementary and middle schools contributed 136 student assignments. The study demonstrated use of this methodology within an evaluation of a complex urban reform initiative. Inferences from the data were analyzed for their technical quality and usefulness. Overall, the technical quality of the approach was reasonable, but anchor papers have been selected and the rubric refined to improve future generalizability. The indicators show promise for use in school or district self-evaluation efforts, not only in monitoring progress, but in identifying areas for administrative attention, professional development, and teacher reflection. Appendixes contain the elementary student work/teacher assignment notebook, a language arts scoring rubric, the teacher interview protocol, and class assignment paper summaries. (Contains 19 tables, 5 figures, and 28 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

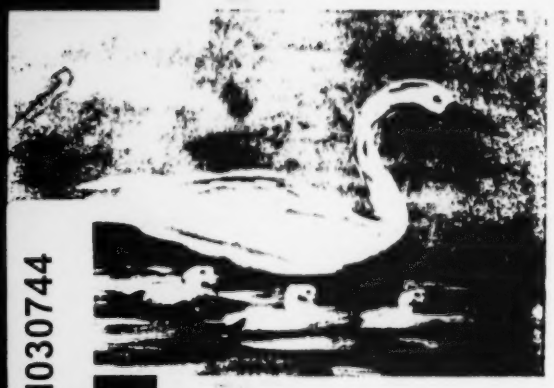
ED 440 132

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Developing Indicators of Classroom Practice to Monitor and Support School Reform

CSE Technical Report 513

Pamela R. Aschbacher
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)/UCLA



TM030744

UCLA Center for the Study of Evaluation

In Collaboration With

BEST COPY AVAILABLE

Developing Indicators of Classroom Practice to Monitor and Support School Reform

CSE Technical Report 513

Pamela R. Aschbacher
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)/UCLA

December 1999

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 1.2 Assessment in Action Pamela R. Aschbacher, Project Director CRESST/UCLA

Copyright © 1999 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

DEVELOPING INDICATORS OF CLASSROOM PRACTICE TO MONITOR AND SUPPORT SCHOOL REFORM[1]

Pamela R. Aschbacher
CRESST/University of California, Los Angeles

Abstract

This report describes the development of indicators of classroom practice for monitoring and improving the quality of school reforms. The work entailed development of a rubric to rate key facets of classroom practice based on assignments and samples of student work. This approach was used to describe the intellectual challenge of class assignments, the alignment of tasks with learning goals and grading criteria, clarity of criteria for success, and provision of informative feedback to students. It also compared teacher judgments of student work with external rater judgments using a school district's standards-based rubric. The study demonstrated use of this methodology within an evaluation of a complex urban reform initiative. Inferences from the data were analyzed for their technical quality and usefulness. Overall, the technical quality of the approach was reasonable, but anchor papers have been selected and the rubric refined to improve future generalizability. The indicators show promise for use in school or district self-evaluation efforts, not only in monitoring progress but in identifying areas for administrative attention, professional development, and teacher reflection.

Introduction

There is a well-known truism in education: The heart of school reform is what happens in the classroom. Unfortunately, although many millions of dollars have been spent to improve what goes on there and what students learn as a result, we do not yet have efficient and effective ways to monitor classroom practice. Evaluation of educational reforms has typically relied on some combination of methods such as

[1] I would like to offer special thanks to several senior colleagues at CRESST and the wonderful team of research assistants who collaborated wholeheartedly with me on this project: to CRESST's co-directors Eva Baker for suggesting the "quest" for classroom indicators and Bob Linn for his very generous help with the generalizability analyses; to Lindsay Clare and Joan Herman for their deep insights and unfailing support; to Joan Rector Steinberg, Jenny Pascal, Rosa Valdés, Roy Zimmerman, and Diane Alvarez for their tremendous efforts to obtain data, develop rubrics, rate assignments and student work, and identify anchor papers; to Rosa and to Xiaoxia Ai for their help with data analyses; and to Joanne Michiuye for her incredibly efficient and cheerful administrative assistance. We all worked hard, shared good times and bad, and learned a lot from each other. I would also like to thank the two dozen teachers who shared their teaching with us so that we might learn from them and try to offer some guidance to others.

observations of classrooms, teacher surveys, and interviews. All of these tend to be complex and labor-intensive and may provide only limited or biased views of actual practice. Certainly they are not very conducive to routine, large-scale use. The goal of the work reported here was to develop efficient indicators of classroom practice that will not only monitor reform efforts but also support the improvement of teaching and learning by focusing attention on critical aspects of practice. Set within an evaluation of the Los Angeles Annenberg Metropolitan Project (LAAMP), this study is part of a strand of research at CRESST on the design of effective indicator systems.

What are educational "indicators?" They are statistics that typically measure some aspects of desired educational outcomes or describe essential features of the education system. They are meant to be used by policymakers and others to assess how a school, district, state, or the nation is doing against a standard, over time, or in comparison with others (Oakes, 1986). Typical indicators include student achievement test scores, dropout rates, graduation rates, and course-taking patterns. They may also include teacher experience and preparation, curriculum topics covered at particular grades, and so forth.

There has been considerable interest in educational indicators in the past two decades, although their use dates back to the middle of the last century. When the first U.S. Department of Education was established in 1867, it was charged with collecting and publishing annual statistics to monitor the condition and progress of education. In response to the widespread criticism of public education over the last two decades since reports such as *A Nation At Risk* (National Commission on Excellence in Education, 1983), policymakers have focused increased attention on educational indicators to monitor the state of education and to motivate improvement. During the 1980s, according to Smith (1988), nearly all the major national or state education groups or agencies were involved in indicators. For example, the National Center for Education Statistics (NCES) of the U.S. Department of Education revised its annual report on the "condition of education" to focus on indicators and published the Secretary of Education's "wall chart" that compared states by means of a variety of educational indicators. In addition, the Council of Chief State School Officers adopted a resolution calling for a national system of standardized indicators. Since then, many states have begun using their own indicators to monitor education reforms, and with the decentralization of much

education funding and calls for school accountability, districts have joined states in using indicators to monitor the "health" of schools.

Indicators are of particular interest for at least two reasons. They can provide a consistent measure across a wide variety of types of programs, but they are more than a mere measurement tool. Since indicators can direct attention toward certain facets of the education system and away from others, they can have a powerful impact on what happens, as noted by Porter a decade ago (1988):

> Indicators could become more than just objective data about the health of the education system; they could become the working definitions of what constitutes good health. (p. 505)

Hence, decisions about indicators—what to measure, who determines it, and how to make sense of the data—have the potential for very significant effects on education. In a quandary about what to do and realizing the stakes involved, many states and districts have looked to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and other experts over the past few years for help in designing indicators. CRESST has recommended development of comprehensive *systems* of indicators (Baker & Linn, 1998) and has initiated a number of research studies in this area. The research reported here was conceived in this context.

The ability to provide consistent measures across a variety of program types is very desirable in the current era of school reform. During the same period of increasing interest in indicators, the nature of typical school reform programs evolved from small, often subject-matter-focused efforts towards large-scale, systemic, comprehensive programs, such as New American Schools, the Annenberg Challenge, and Title I. Such reforms are intended to be comprehensive yet flexible to accommodate the unique needs of the many individual schools or districts involved. The resultant variation within such programs, however, provides a serious challenge for monitoring progress and evaluating the results of these huge investments of human and capital resources. Baker and Linn (1998) have suggested that comprehensive indicator systems could address this concern. Such systems could include not only measures of student outcomes, such as test scores and graduation rates, but also measures tied to specific goals of the reform, such as parental involvement and professional development (Los Angeles Compact on Evaluation, 1998).

A common feature of many of these complex reforms is their call for the development of school and district capacity for self-evaluation. As schools and districts struggle to develop action plans to improve teaching and learning, they need simple, effective methods for collecting and utilizing data on how well they are doing. As one principal exclaimed in an interview last year (Aschbacher, 1998):

> Can't someone design some kind of measuring tool to measure progress, other than district and state test scores? How do we measure what is going on in the classroom?!

Although much of the push for indicators has come from policymakers, the quote above illustrates that school professionals obviously want to know not only how well students are doing over time or compared to other schools and districts, but *why* (Richards, 1988). To provide some explanatory power, a number of researchers and policymakers have asserted the importance of measuring not just what students have learned but what they have had the *opportunity* to learn (Carey, 1989; National Council on Education Standards and Testing, 1992; Oakes & Carey 1989; Selden, 1988; Shavelson, McDonnell, & Oakes, 1989). We agree with David (1988), who claimed there is significant constructive potential for educational improvement based on local indicators that capture what happens in classrooms and that focus on the quality of practice.

As many researchers and others have noted, whatever is measured tends to take on heightened importance, or as H. D. Hoover wittily captured the notion: WYTIWYG —what you test is what you get (1996). Thus it is wise to select things to measure that are truly worth focusing on. Our experience in several studies at CRESST involving professional development of teachers suggested possible areas of focus for classroom indicators. In these studies, many teachers experienced difficulties in maintaining high standards for student achievement and in developing learning and assessment activities and grading criteria aligned with student standards (Aschbacher, 1994; Aschbacher & Herman, 1991; Aschbacher & Rector, 1996). Teachers' curriculum and instruction decisions tended to be driven by activities rather than by desired student outcomes, and the activities often emphasized participation rather than rigorous thinking or use of content knowledge. It was our intent that indicators linked to well-established features of good instruction (complex thinking and use of content knowledge; coherent alignment of goals, tasks, and criteria; clear targets for success; and informative feedback) could help describe the quality of learning opportunities afforded to

students as well as guide teachers' attention toward these areas to improve teaching and learning.

An evaluation conducted by the author several years ago demonstrated the feasibility and value of assessing the quality of classroom assignments along with the student work that was elicited by them (Aschbacher, 1992). By examining student portfolios, including both the student work and the assignments to which students responded, we found that students were more likely to attain program goals (e.g., to learn to make interdisciplinary connections) when their assignments were specifically designed to elicit the desired kind of thinking. Newmann and Weglage (1995) also rated the quality of teacher assignments (in math and social studies) and linked this to the quality of student work. A version of their approach is currently being used in the evaluation of the Annenberg Challenge in Chicago (Newmann, Lopez, & Bryk, 1998).

In the New American Schools model developed in Los Angeles, known as the Los Angeles Learning Centers, in the Critical Friends Groups promoted by the Annenberg Institute, and in many other reform efforts of the past decade or so, teachers have begun to come together to reflect on student work. Unfortunately there have been few guidelines to shape their conversations and help teachers see the connections among expectations for student learning, assignments given, criteria used to provide feedback and to grade students, and the nature of the resulting student achievement.[2] The goal of our work is to support teachers' reflective practice by focusing attention on critical dimensions of good teaching potentially under their control and on the consequences for student achievement.

This work follows in the footsteps of previous CRESST work on generic models for the development of performance assessments in several subject areas (Baker, Aschbacher, Niemi, & Sato, 1992; Baker, Freeman & Clayton, 1991). Our strategy of focusing on generic aspects of strong practice, which are relevant across a broad array of subject area reforms, is intended to facilitate teachers' improved practice (Baker, 1997).

Our work is intended to provide two valuable tools to help schools and districts enhance their capacity for improving education. The first is a set of indicators of classroom practice that provide an alternative to observation and

---

[2] A new resource is now available from Harvard's Project Zero and the Annenberg Institute for School Reform: *Looking together at student work: A companion guide to assessing student learning* by Blyth, Allen, and Powell (1999).

teacher self-report. Such indicators could be used in research on teaching and learning, in large-scale evaluations, and in local self-evaluation efforts to monitor progress in instructional quality along with student performance over time. The second tool is a rubric with guidelines for describing the nature and quality of classroom assignments and linking them to student work. Just as looking at student work has become a popular and effective strategy for encouraging teachers to be more reflective about their classrooms, the rubric is meant to deepen and extend teachers' reflections on the quality of an assignment and its impact on the nature of student work. Such a tool should be useful in both pre-service and in-service professional development.

Work towards these goals is progressing in stages. The first stage, reported here, includes a number of steps: initial development of the specifications for the assignments and student work to be collected, collection of the first data, drafting a rubric for evaluating particular characteristics of assignments, training raters and applying the rubric, analyzing results of the ratings and making comparisons of inferences from ratings to those from teacher interviews, compilation of anchor assignments to illustrate application of the rubric, and revision of strategies and instruments. Another report (Clare & Valdés, 1999) analyzes the relationship between this approach and classroom observations. The second stage, now in progress, includes applying the revised rubrics to new data collected during the 1998-99 academic year, analyzing those data, making comparisons with previous data and with both interviews and classroom observations, and making revisions. A third stage will entail field trials in which one or more schools adapt this approach for their own self-evaluations.

Since our focus here was on the development of a new methodology, the research questions addressed in this study concerned its technical quality and usefulness as outlined below:

Technical quality

- How reliable were the ratings of assignments?
- How independent were the rating dimensions?
- How consistent were teachers' assignments?
- Did ratings of assignments and interviews provide similar estimates of practice?

Usefulness

1. What can this methodology tell us about the classroom learning environment?

   - Are students intellectually challenged?
   - Are students given clear criteria for success?
   - Are students given informative feedback?
   - Are students given "coherent" assignments—tasks aligned with learning goals and criteria?
   - How do teachers perceive student performance?

2. What can this methodology tell us about the relationship between the learning environment and student achievement?

3. Are learning environments equitable for all students?

4. Did teachers' reflections on assignments prove useful to them?

Method

Our general approach to this program of research and development was to respect the evaluation context in which our work was situated and yet to strive to develop tools that might work well in a broad range of instructional settings. For example, we selected language arts as the target curriculum because increased literacy was a primary goal for every LAAMP school, and we selected elementary and middle school grades in which to work in part because LAAMP efforts were directed primarily at those levels rather than high school. In an effort to develop fairly generic tools, we selected two different grade levels in which to work, and we created a menu of fairly generic language arts assignments that could be considered typical in many different language arts classes.

We utilized data from several sources: a sample of teachers' assignments in language arts at Grades 3 and 7 along with samples of high- and medium-level student work elicited by those assignments; teachers' contextual descriptions of their assignments, including their learning goals and criteria for judging student work; interviews with teachers about one of the assignments and related student work; and general background information on the teachers and their classes.

Participants

Twenty-four teachers from eight LAAMP schools (12 teachers from four elementary schools and 12 from four middle schools) participated in the study and contributed 136 assignments, with four pieces of student work for each assignment. Middle school teachers submitted assignments and student work from just one of their classes. We had requested language arts assignments and student work from 4 teachers per school, for a total of 32 teachers (i.e., classes). This sampling plan was designed to include most of the teachers on track at the time of data collection, both bilingual and English-only instruction, and a range of teacher experience, classroom practices, and student achievement. Teachers participated voluntarily, and their principals had to give permission as well. The overall participation rate was 75%. Teachers received a stipend of $100 for their efforts beyond the normal school day activities to compile the requested data.

Data collection was focused at two grades (third and seventh) to explore the feasibility of this approach in both elementary and secondary settings. The choice of third and seventh grades was based on the likely availability of student performance assessment data at those grades in the future, which would be useful in attempts to validate classroom indicators of complex learning opportunities. In addition, third grade is a pivotal year in literacy, reflecting early efforts at reading and writing instruction and the readying of students to begin work in the disciplines. Seventh grade represents the center of middle school efforts.

The sample of assignments and student work actually submitted for review by teachers represented a broader range of grades than researchers requested. Three of the elementary schools had combined grades within classrooms, such as second and third grades together and third- and fourth-grade combinations, so it was not possible to obtain assignments from one grade alone in these schools. In addition, virtually all the third-grade teachers at one school were new, emergency credentialed teachers, and their principal did not allow them to participate. At that school, work was submitted primarily from second-grade classes. In the middle schools, some of the seventh-grade language arts teachers were unable or unwilling to participate during spring 1998 whereas some sixth- and eighth-grade teachers were eager to participate. Thus about half the middle school assignments submitted were from seventh grade, a quarter from sixth grade, and a quarter from eighth grade. Descriptions of the assignments below refer to "elementary" and "middle school" assignments because they were not gathered exclusively from third and

seventh grades. For analysis of student work, however, a sample including only third- and seventh-grade writing was drawn.

Procedures

In the early spring of the year, each teacher received a binder of materials that included

- a cover letter describing the purpose of the study;
- a consent form and stipend information;
- directions for assembling assignments and student work samples and how and when to submit these materials;
- a color-coded cover sheet for each type of assignment, with space for teachers to describe the assignment, learning goals, assessment strategies and criteria, and range of student performance on the assignment;
- a one-page survey of teacher background and classroom context (e.g., years of teaching experience, class size, and student English fluency); and
- preprinted identification code labels for teacher and student work to maintain confidentiality of the data.

Teachers were informed of the general purposes of the study, to examine the nature of school improvement for the Annenberg Challenge in Los Angeles. Since the rubric for rating assignments was still being developed at the time of this first stage of data collection, teachers were not told about the specific criteria by which their assignments and student work would be analyzed. (See Appendix A for sample teacher notebook.)

The sample of assignments and student work was designed to provide a broad picture of language arts curriculum and instruction without overburdening teachers. The sample asked for assignments that might reveal changes in curricular rigor over time—such as various types of writing assignments and a major challenging project. Each teacher was asked to submit a sample of six assignments from the spring of 1998:

1. one reading comprehension assignment,
2. one writing assignment with a draft,

3. one writing assignment in a content area (asked of elementary teachers only since middle school English teachers could not be expected to use such an assignment routinely),

4. one challenging, major project with a written component (two such assignments requested at the middle school level to compensate for the lack of writing in a content area), and

5. two typical homework assignments.

Teachers were asked to submit four samples of student work (two for "high" -level achievement and two for "moderate" -level achievement) for each of these assignments. This sample was designed to provide some insight into teachers' expectations for student learning and performance as well as illustrative examples of the types of student performance elicited in these classrooms.

In late spring, 10 teachers (4 elementary, 6 middle school) were interviewed by researchers in depth for about an hour about one of their assignments and the related student work. These interviews were to serve two purposes: (a) to provide additional information to help evaluators more fully describe the learning opportunities and expectations that students were afforded and the kind of work they produced in response; and (b) to help validate the inferences made from the submitted written documents and determine the feasibility and validity of a possible "by mail alone" data collection strategy. Researchers audiotaped the interviews with teachers' permission and transcribed them for analysis.

Measures

To describe the nature of classroom assignments, researchers developed a rubric based on results of past CRESST research and evaluation studies of teaching practices in a variety of school reform efforts as noted above. Researchers first examined a range of typical language arts assignments for elementary and middle school, identified a number of potential variables that might distinguish stronger from weaker instructional settings, applied rudimentary scales to several assignments at each grade level, discussed the results, identified the most promising variables, refined rubric definitions for each scale, reapplied them to a sample of assignments, and revised them as needed. Finally, that draft of the rubric was used to rate the 136 assignments submitted for this study. (See Appendix B for draft rubrics.)

The rubric used here consisted of

- six descriptive scales
  - type of assignment
  - type of content knowledge used
  - type of student response
  - type of choice students were given,
  - grading dimensions used
  - types of feedback provided
- five 4-point evaluative scales
  - cognitive demands of the task
  - clarity of grading
  - alignment of task with learning goals
  - alignment of grading criteria with learning goals
  - overall task quality

Because virtually none of the assignments involved the use of technology, we did not develop a scale for this aspect of the assignments, contrary to our original plans. Although the improvement plans of LAAMP School Families[3] called for use of technology, it simply had not been widely implemented at the time of this data collection.

Each assignment was rated by two trained raters, who were CRESST researchers with teaching experience. Four different raters participated in scoring the work. Raters scored elementary and middle school assignments separately. Within each level, all assignments regardless of type were rated in random order. The average percent of exact agreement between two raters across five evaluative scales for five types of assignments was 53.5%; the average plus-or-minus-one-point agreement between two raters was 99.7%. Details about interrater reliability are provided in the Results section. Analyses involving ratings of assignments utilized the average score for the two raters since there was not 100% exact agreement.

The students' final written work for the writing-with-a-draft assignment was rated by two bilingual raters with teaching experience using three standards-based,

---

[3] "School Family" is the term used in LAAMP for a set of elementary, middle, and one or more high schools, typically in a feeder pattern, that develop a joint action plan and work together on common goals and strategies for improvement.

4-point writing scales (Organization, Content, and MUGS[4]) from the recent joint LAUSD, CRESST, and UTLA Language Arts Project (LAP rubric; see Higuchi, 1996). We did not rate work done by students outside the targeted third and seventh grades, nor work on one elementary assignment that was simply too unclear to score fairly. There were 16 elementary essays in Spanish, 16 elementary essays in English, and 24 middle school essays in English.

We rated separately the student work written in Spanish from that written in English. Unfortunately no benchmark papers for the LAP scales were available in Spanish to guide raters, and our bilingual raters failed to reach sufficient agreement within the time available to include the Spanish essays in further analyses for this study. Interrater correlations on the three LAP scales applied to third-grade writing in Spanish ranged from .24 to .43; exact agreement on these 4-point scales ranged from 25% to 37%; one-point agreement ranged from 81% to 94%.

For the student work in English, the average percent exact agreement between raters across the three scales was 56%; the average one-point agreement was 92%. Although one-point agreement between raters was about the same for each of the three scales (92-93%), the exact agreement was much higher on MUGS (69%) and Organization (60%) than on the Content scale (38%).

Although the amount of student work analyzed here was quite small, and the interrater reliability was not as high as one might like, we used these data to conduct further analyses, reported below, to illustrate the value of relating student performance to characteristics of classroom practice.

Follow-up interview questions for teachers addressed such issues as how the assignment was related to prior and subsequent instruction; the learning goals addressed in the assignment; alignment of learning goals, grading criteria, and district or state standards; the teacher's expectations for student work; and how the teacher used information on student performance in these assignments (e.g., for revising instruction, placing students, planning remediation, and so forth). (See Appendix C for Teacher Interview Protocol.)

---

[4] MUGS is an acronym for a very common set of criteria for judging language arts work: mechanics, usage, grammar, and spelling.

Results

In the development of potential new indicators, two characteristics are crucial: their technical quality and their usefulness. This paper explores four aspects of technical quality: reliability of assignment ratings, independence of scales, consistency of assignment types, and validity of ratings compared to interview data. Aspects of utility addressed here include the capacity to describe practice and its relationship to student achievement and other variables. Results for each of the research questions outlined in the introduction are discussed under these two major headings below.

Technical Quality

The overall technical quality of our approach to measuring classroom practice through ratings of assignments was reasonably good for this first stage of the development process. Interrater agreement on the descriptive scales was high; however, interrater reliability on the five evaluative scales was only moderate, ranging from .53 to .74, and therefore needs improvement. We have already begun refinement of rating scale definitions and establishment of anchor assignments for many of the points for each evaluative scale (see Appendix D). Of the five evaluative scales, two pairs were moderately highly correlated (about .65 and .74). If these results hold for analysis of the next data set, two of these scales could probably be dropped eventually, thereby streamlining the method. Generalizability analyses revealed that it is desirable to sample at least three or four different types of assignments, because there are differences in mean scores among assignments. Ratings of assignments generally agreed with holistic estimates of assignment quality based on interviews, but interviews provided far more detail.

How reliable were assignment ratings? Interrater consistency or reliability is a fundamental feature of any measurement tool because valid inferences cannot be made if trained raters disagree about the "value" of an assignment. The goal of high interrater reliability, however, was a considerable challenge in this study for several reasons. First, this study was the initial application of new rubrics with no previously agreed upon anchor papers to guide raters. In addition, reliability was a challenge because the scoring of assignments required a rather complex analysis of materials. Not only were raters supposed to evaluate what amounts to several performance assessments for teachers, but the evidence to be reviewed in each case was not a simple essay, as is often true with student performance assessments, but a

combination of as many as four types of documents: the cover sheet descriptions of their assignments completed by teachers, the task directions for students that some teachers submitted, any rubrics or grading guidelines they may have submitted, and four samples of student work. In some cases, teachers' task descriptions were minimal, and it was necessary for a rater to look at the student work to clarify what the task actually entailed. Further challenging the attainment of rater reliability were the number of raters who participated (four) and the wide variation in the types of content they encountered (five different types of assignments at two different grade levels).

1. Evaluative scales. We examined interrater reliability on the five evaluative scales developed in this study using two methods: Spearman-Brown correlations and percent agreement between raters ("exact" as well as "plus-or-minus-one-point" agreement).[5] Table 1 displays the interrater reliability coefficients for the five evaluative scales used to assess classroom assignments. Table 2 displays the percent agreement consistency across raters. Note that in both cases, the reliabilities were based on all five types of assignments combined.

Tables 1 and 2 reveal that raters tended to agree with each other but not often enough or closely enough for this first version of the rubric to be used again in future without revision, anchor papers, and additional training. For example, correlation coefficients ranged from a low of .53 to a high of .82, with the majority of the correlations under .80. Raters were nearly always within one point of each other (from 91.2% to 100% of the time), but raters agreed exactly far less often (from 47.1% to 60.8% of the time). Both agreement and correlation coefficients varied quite a bit by scale and by grade level, as shown in the tables.

For example, the interrater correlation for Grading Clarity was .82 for middle school assignments but only .62 for elementary assignments. The Overall Quality scale had the lowest interrater reliability (approximately .53 for both elementary and secondary assignments). These two different types of interrater reliability, however, did not yield a common pattern of results: The scales with the higher reliability coefficients did not have higher percentages of rater agreement.

---

[5] Exact agreement is the percent of cases in which one rater awards exactly the same score as the second rater; in plus-or-minus-one-point agreement, the first rater awards a score that is not more than one point higher or lower than the score given by the second rater.

Table 1

Interrater Reliability Coefficients for Five Scales Evaluating All Classroom Assignments

Scales	Elementary assignments (n = 86)	Middle school assignments (n = 50)	All grades (n = 136)
Cognitive demands	.68	.54	.61
Grading clarity	.62	.82	.73
Alignment of learning goals and task	.75	.66	.72
Alignment of learning goals and grading	.82	.64	.74
Overall task quality	.53	.54	.53
Average overall	.68	.64	.67

Table 2

Interrater Percent Agreement for Five Scales Evaluating All Classroom Assignments

Scales	Elementary assignments (n = 86)	Middle school assignments (n = 50)	All grades (n = 136)
Cognitive demands	51.5 (92.6)	47.1 (98.5)	52.2 (99.6)
Grading clarity	57.4 (92.6)	66.2 (100.0)	60.8 (100.0)
Alignment of learning goals and task	64.7 (92.7)	48.5 (98.5)	57.4 (99.6)
Alignment of learning goals and grading	50.0 (91.2)	54.4 (97.1)	52.8 (97.8)
Overall task quality	48.5 (92.6)	54.4 (100.0)	52.6 (100.0)
Average overall	54.4 (92.4)	54.1 (98.8)	53.5 (99.7)

Note. Percent exact agreement is given first; plus-or-minus-one agreement is in parentheses.

These moderate interrater reliabilities suggest that this first version of the rubric needs tighter definitions and clear anchor papers for training to reduce the variation among raters. Because raters were within one point of each other so often, this should be possible. In addition, future raters should have more extensive training, with specialization by grade level and possibly by type of assignment. On these 4-point scales, it would be highly desirable to achieve significantly better exact agreement (perhaps 80% or better) for the rubric to be helpful to teachers in improving practice or to provide reliable indicators of classroom practice.

2. Descriptive scales. Six descriptive scales were also used in this study, the most relevant and promising of which were (a) the type of content knowledge the student would have to use in the task and (b) the type of feedback provided by the

teacher. Because these scales consisted of categories with no ordinal meaning, we calculated interrater agreement by simply counting the number of times raters disagreed on the categories they selected to describe each assignment. Exact agreement was extremely high, 98% to 99%. Raters disagreed on the categories of content knowledge only 3 times out of 136 assignments, and disagreed only once for type of feedback.

How independent were the rating dimensions? Monitoring progress in large-scale settings puts a premium on efficiency—in terms of both costs and the time it takes to score and report back the results—so we examined ways to streamline our method. In this approach, both collecting assignments from teachers and rating them are labor intensive activities; thus, it is desirable to use as few dimensions or rating scales as possible. We calculated correlations among all scales to see whether some of them might be so highly correlated that one or more could be omitted as redundant. Tables 3, 4, and 5 present these correlations for elementary assignments, middle assignments, and both grades combined. In each table, the correlations are among all five rating scales, where each scale is applied to all types of assignments.

Table 3

Correlations Among Rating Scales for Elementary School Assignments

	Grading clarity	Goals/Task	Goals/Grading	Overall quality
Cognitive demands	.14	.16	.24	.73
Grading clarity		.32	.67	.24
Goals/Task			.41	.47
Goals/Grading				.37

Table 4

Correlations Among Rating Scales for Middle School Assignments

	Grading clarity	Goals/Task	Goals/Grading	Overall quality
Cognitive demands	.31	.36	.33	.75
Grading clarity		.26	.67	.42
Goals/Task			.40	.58
Goals/Grading				.52

Table 5

Correlations Among Rating Scales for All Assignments

	Grading clarity	Goals/Task	Goals/Grading	Overall quality
Cognitive demands	.23	.20	.27	.74
Grading clarity		.26	.65	.32
Goals/Task			.39	.49
Goals/Grading				.42

As Tables 3, 4, and 5 reveal, two pairs of scales had consistently high intercorrelations: the Overall Quality scale with the Cognitive Demands scale (.74), and the Clarity of Grading scale with the Alignment of Grading With Learning Goals scale (.65). Neither case is surprising because within each pair the scales are related by definition. We defined tasks with high Overall Quality as those that challenge students to use complex thinking (i.e., high Cognitive Demands) as well as demonstrate other features such as coherence of goals, task, and grading. Likewise, the Alignment of Grading With Learning Goals scale depends to a great extent on the degree of clarity of the grading expectations.

Although it is desirable to reduce redundancy, it is difficult to select one scale in each pair over the other at this point. In each pair, referring back to Tables 1 and 2, neither scale is much more reliable than the other, although Cognitive Demands has slightly higher interrater correlations than does Overall Task Quality. Because teachers typically have difficulty articulating their learning goals for students, and two scales (Alignment of Goals With Task, and Alignment of Goals With Grading) measure aspects of this problem, we have decided to attempt to define a new scale, Clarity and Elaboration of Goals, and to determine, through factor analyses of new data from the 1998/99 academic year, which of these scales is most reliable, independent, and useful. (See also the generalizability studies reported below.)

How consistent were teachers' assignments? We collected a variety of assignments in this study: six assignments from each classroom/teacher, which represented five different *types* of assignments. Could fewer assignments be collected and still provide a reasonable description of the practice in a given classroom? If so, which types of assignment might be the most useful to collect? The amount of data needed is a function of how consistent teachers tend to be across the various learning activities they use in their classes. The more each assignment is like another within a class, the fewer assignments need to be sampled to have a good

estimate of the type of learning environment there. It would also cut costs to reduce the number of raters needed to score the assignments.

To address these concerns, we conducted generalizability studies to investigate the consistency of teachers' assignments across the five types of assignments we collected. We analyzed elementary and middle school teachers together, using the 19 teachers for whom we had complete data for all 6 assignments, with ratings from 2 raters on each of the 5 evaluative dimensions.

We computed error variances for (a) relative decisions, called Var(d1) or Var(d2), and (b) absolute decisions, called Var(D1) and Var(D2), where different teachers might be rated by different raters and have different assignments. The first of these error variances, (d1) and (D1), were calculated with dimensions as a random factor. That is, the generalization is across dimensions as well as assignments and raters. The second one in each pair, (d2) and (D2), treats dimensions as a fixed factor (i.e., these dimensions are the ones we care about, not the larger universe of possible dimensions) and gives the average error for a single dimension.

With dimensions fixed, the results look pretty good (Table 6). The teacher score variance of .079 is considerably larger than the error variance for absolute decisions (.019; how well a teacher can do against a criterion, not relative to other teachers). Consequently, with 2 raters and 6 assignments, we get a dependability coefficient of .806. A reliability of .8 is reasonably good for the number of separate pieces of information we have about a teacher (six assignments).

Some other things of interest relate to the individual variance components (VC). The VC of .0056 for one rater (.0028 for two) shows that our training has been relatively effective in avoiding large differences between raters in their leniency-stringency of rating. The VC of .0111 for one assignment (.00185 for 6 assignments), however, indicates that there are differences in mean scores among assignments that make it important to average those out over several assignments (as we have to a fair degree with six). The VC of .109 for the teacher by assignment interaction (.00225 with 6 assignments) also says that it is important to have multiple assignments per teacher. The same could be said about the TARD, error component. The VC of .2366 for one dimension (compared to .04732 for 5 dimensions) reveals the value of using several dimensions to rate assignments.

Table 6

Summary of Results of Generalizability and Dependability Studies on Assignments, Dimensions, and Raters

Effect	Variance component	Variance component(Des)
Teacher	0.0792	0.0792
Assignment	0.0111	0.00185
Rater	0.0056	0.0028
Dimension	0.2366	0.04732
TA	0.109	0.00225
TR	0.0135	0.00675
TD	0.0591	0.01182
AR	0	0
AD	0.0021	0.00007
RD	0	0
TAR	0.0328	0.002733
TAD	0.2276	0.007587
TRD	0.02282	0.002282
ARD	0	0
TARD,error	0.16	0.002667
Var(d1)	0.62482	0.036089
Var(d2)	0.3153	0.0144
Var(D1)	0.88022	0.088129
Var(D2)	0.332	0.01905
G-Coe(1)	0.112497	0.686971
G-Coe(2)	0.20076	0.846154
D-Coe(1)	0.08255	0.47332
D-Coe(2)	0.192607	0.806107

Next we computed G-study results for six different designs (i.e., from 3 to 6 assignments and 1 to 2 raters) to determine whether in the future we could streamline the design. The results of greatest interest are those for G-Coe(2), the generalizability coefficients for a fixed dimension where teachers have the same raters and assignments (see Table 7). The G-coefficients for 4 assignments and 2 raters (.81) and for 3 assignments with 2 raters (.78), as noted in the table, indicate that both of these are reasonable designs. None of the designs with one rater have sufficiently high coefficients (anywhere near .8) to support their use at this point. These findings suggest using two raters to rate at least three assignments on all

Table 7

G-Study Results for Different Possible Designs (Numbers of Assignments and Raters)

Design:	A=3; R=1	A=3; R=2	A=4; R=1	A=4; R=2	A=6; R=1	A=6; R=2
G-Coe (2):	.6667	.7822	.7054	.8129	.7489	.8461

Note. A = assignments; R = raters.

dimensions for the next study in this series. Better rater training, including anchor papers for most points of all the scale dimensions, should help further.

Did ratings of assignments and interviews provide similar estimates of classroom practice? One goal of this research was to explore the extent to which ratings of assignments might serve as a proxy for descriptions of classroom practice derived from other methods such as teacher interviews. We used two strategies to shed some light on this objective:

1. comparing overall estimates of the quality of practice based on the interview alone to ratings based on the assignment materials submitted;

2. comparing teachers' answers to questions that appeared on the interview with those from the cover sheet submitted with assignments.

Our general conclusion was that ratings of assignments generally agreed with a holistic estimate of assignment quality based on the interview, but that the interview provided far more detail, as expected. Based on responses in both settings, many teachers appeared to have somewhat vague and/or fluid notions of what learning goals they pursued and what criteria were important for evaluating student performance. Interviewers had the advantage over raters of being able to probe when a teacher's response was vague. Raters, on the other hand, were forced to deal with vague information and could have drawn different inferences (about alignment of goals, tasks, and criteria, for example) than interviewers with greater information. It appeared that even though teachers had received a stipend for submitting materials, most of them put together their notebooks quickly and did not make extensive comments on the cover sheets. To some extent, this may have been a function of collecting data during April to June, a period in which teachers often seem tired and less engaged in activities that are not a high priority for them.

Our first comparison was between the interview data and the ratings of the writing assignment on which the interview focused. We compared one researcher's

holistic estimate of the overall quality of practice based on the interview alone[6] with the sum of two researchers' ratings of the assignment based on the written materials submitted by the teacher. Table 8 shows the holistic interview scores compared to the sum of two raters' scores on the five evaluative scales for each teacher interviewed.

The results suggest that although there was not an exact correspondence with these two sets of ratings, the holistic interview score was reasonably aligned with our rubric-based ratings of one assignment. The two judgments disagreed most on Teachers 1, 4, and 8. In two cases, the ratings were lower than expected from the holistic score; the third case was in the opposite direction.

In the second comparison, we focused on two areas that appeared in both the interview and cover sheet and that figured in our ratings of assignment quality: the learning goals and the grading criteria. For each area, we compared what a teacher said on the cover sheet with what he or she said in the interview.

We found that 9 out of 10 teachers described their learning goals slightly differently in the interview compared to the cover sheet, and 8 teachers described their grading criteria somewhat differently. It is not clear that one source of data is

Table 8

Holistic Interview Scores Compared to Ratings of Assignment Quality

Teacher	Holistic interview score (1-5 scale)	Ratings of assignment (10-40 scale)
T1	1	24
T2	1.5	22
T3	2	22
T4	2.5	20
T5	2.5	23
T6	3	26
T7	3	32
T8	3.5	22
T9	4	36
T10	4.5	39

[6] This was in informal rating of the interview data by one researcher who had not participated in the actual interviews and had not met the teachers. It used a 1-to-5 scale, where 1 reflected very weak teaching; 5 reflected very strong, coherent, challenging teaching practices.

more "accurate" than the other. It is possible that teachers' views of why they had students do certain assignments and what they hoped to see in student performances could have changed from the time and setting in which they completed the notebook of written materials to the time and setting in which they were interviewed. Teachers were typically terse on the cover sheet, so the discrepancies might also reflect a lack of attention to detail on the written materials rather than true differences. Nonetheless, results suggest that ratings of the alignment of goals, tasks, and criteria based on written materials alone might in some cases be affected by teachers' lack of precision or care in completing the forms. This is analogous to concerns over whether high school student test scores reflect poor understanding and/or poor motivation to demonstrate what they know. We did not emphasize in our directions to teachers that they would be judged on the basis of the words they used to describe their practice. The stipend was evidently an incentive only to participate, not to complete the forms with great care. In future, the directions for teachers should be refined to assure greater motivation to express themselves carefully and accurately.

Figure 1 illustrates some of the different ways that teachers described their goals and criteria from the cover sheet to the interview. In general, they tended to include some elements in one place that were not mentioned in the other (as highlighted in the figure). Teachers seldom defined their terms, so it was sometimes hard to judge whether they meant the same things from cover sheet to interview (e.g., for Teacher 6, does "clarity and style" mean "interesting, unique, clever, well organized . . . painting a picture"?).

Some teachers seemed uncertain about their learning goals for a particular assignment or did not develop their criteria prior to assigning the task to students (see Teacher #4 above). When teachers have an amorphous sense of what they want students to be able to do, they may mention certain elements of criteria or goals on the cover sheet but include different elements in their interview. Neither is necessarily "inaccurate." This can occur even when teachers use an elaborate written rubric (which would tend to obtain a high rating on Clarity of Grading Expectations), since they sometimes omit a dimension from the rubric that is actually critical to their stated priorities (see Teacher #5: no rubric dimension for character description).

We also compared the interview and written cover sheet regarding what proportion of their class teachers believed had done well on the assignment. By the

	Written cover sheet	Interview
Goals		
T1	"**Originality**, sentences, paragraphs, writing **creativity**."	"Choose a **character and describe**. Be able to **follow directions**. Be able to **create a story**. Be able to write a draft, **revise it, and go to publish**."
T2	"Read current information; think about it; discuss it. "**Write a letter** to the paper. "**Revision and completion**."	"I wanted the kids to be more aware of violence in the world...to **form their own opinions with supportive facts**, to **take a stand** against violence, etc."
T3	"I am working on the narrative process. I am trying to get students to write in more **detail** about observable things. Also trying to introduce **dialogue**."	"The whole objective of this was for students in their writing process to be very clear on whose **point of view** the story was being told from."
Grading		
T4	"Assessed them on content and written expression."	"How much information they included; whether they understood the brainstorming and find their facts; **how to get information** from books. "I was **not looking at their writing up of it into the paragraph**. "I did not have this criteria in mind ahead of time."
T5	1. "Correctly writing 3 paragraphs (indent, complete sentences, capitalization, punctuation, spelling) with one main idea and details; 2. "Correctly **identifying a character's qualities**; 3. "**Comparing both characters** at final paragraph adequately (not just that they are friends but salient differences)."	"I was looking for mechanics, capitalization, spelling, **not repeating statements, certain vocabulary**. In terms of the analysis, I was looking for an accurate description. I was hoping for **four things** that they could find similar about them in the third paragraph. "I used this rubric. The students and I came up with this 5-point rubric. They would use it first, and then I would check their self-evaluations. 1. "I indented the first sentence in the paragraph 2. "I wrote in complete sentences with capitals at the beginning and periods at the end 3. "I wrote **neatly** 4. "I used correct spelling 5. "I used **interesting words**."
T6	"Rubric: 1. Correct letter heading 2. Introductory paragraph explaining the problem 3. At least 2 causes explained 4. At least 2 solutions explained 5. Closing paragraph 6. Correct language usage 7. Clarity and style" [4 points possible for each, not defined]	"I used a 6-point [sic] scoring guide: 4=very good, 3 is ok, 2 is poor. But I wouldn't define it as a rubric. I attached it... I looked at the introduction paragraph, supporting facts, examples, strong statements, etc.... These two papers are good because they have a lot of details, supportive information, effective use of language, excellent intro, transitions, two solutions, unique, **interesting, well organized, clever**. etc. I look at trying to focus on showing and not telling in writing—**painting a picture**, use of text and **examples**, not taking the reader for granted."

Figure 1. Comparison of teachers' written and interview descriptions of learning goals and of grading criteria.

time of the interview (approximately one to two months after the assignment), teachers tended not to remember how well the class had done on the assignment and typically referred the interviewer to their written cover sheet as more accurate. The interview with one teacher, however, revealed that in fact half the class had not turned in the assignment (according to the teacher this was largely due to their lack of understanding), so the "50%" who did well according to the cover sheet response, was actually only 25% of the class. In future, the cover sheet should be revised to avoid this type of problem.

Not surprisingly, the interview gave a much richer picture than the written cover sheet of the instructional context for the assignment. Teachers tended to describe at length whether the assignment was connected to other subject areas or to other assignments. This provided the interviewer with a better sense of the learning environment, including how organized and well planned the assignment was. While this was not specifically rated, this information could have helped raters understand the nature of the assignment when teachers failed to describe it well on the cover sheet or to provide the directions they had given to students. Raters often had to look at the examples of student work to piece together what students had actually been asked to do. Some teachers also provided elaborate detail in their interviews that clarified whether feedback was provided, when and by whom, how they handled student assessment, and whether or how they worked with student performance standards. This additional detail has the potential to contradict raters' scores based on written materials, such as the descriptive scales about content knowledge required in the task and the type of feedback provided.

Although the above discussion might suggest that interview data were stronger than ratings of written assignments, our experience also revealed a strength of the latter. We discovered that it is critical to look directly at assignments and not rely solely on teachers' descriptions, whether written or oral, because these may have their own bias or inaccuracies. For example, one teacher said on the cover sheet that the goals of his assignment were "test taking skills; writing from an outline." He reiterated these goals throughout the interview. Interviewers and raters alike, prior to inspecting the actual task materials, interpreted this to mean that students were to create a substantive outline for a topic they were going to write about and then write an essay based on that outline. Instead, the "outline" the teacher had referred to was actually a generic set of prompts he had generated that directed students to write a five-paragraph essay:

Introduction: What is the situation to be speculated on?

Body Paragraphs: Speculate about outcomes; base speculation on facts, expert opinions, statistics; demonstrate a logical plan of organization.

Conclusion: Summarize your speculations; leave reader with sense of closure.

Usefulness

This study examined two basic aspects of the utility of our methodology: its usefulness in describing the classroom learning environment, and its usefulness in describing relationships among assignments, student performance, and classroom or teacher characteristics. The rest of this section discusses the results in relation to these two topics.

In general, results of this study suggest that this approach provides very useful information about the classroom learning environment, such as the extent to which students are challenged with significant content and complex thinking in their classroom assignments. This methodology also enabled us to note certain relationships among assignments, teacher and class characteristics, and student outcomes. For example, teachers with more experience at a given grade used assignments of more consistent quality, and higher quality assignments were more often given to classes with higher performing students. In addition, many teachers said that reflecting on their assignments was useful to them.

What can this methodology tell us about the classroom learning environment? As this methodology was used in the LAAMP evaluation, it allowed inferences about a number of important factors in the learning environment of those schools: the extent to which assignments challenged students with significant content and complex thinking; the extent to which students were provided clear criteria for success and feedback to shape their learning; the "coherence"[7] of the learning activities on which students spent their time; and the level of teacher expectations for student work. Figure 2 presents a profile of the learning environment in the elementary and middle school classrooms investigated here as an example of one product of this approach. As the figure notes, a third or less of the reading comprehension, draft writing, and project assignments provided intellectual rigor; one third to one half of the assignments had goals, tasks, and criteria aligned with one another; slightly over one third provided students with clear criteria for

[7] Coherence is used here to mean the extent to which a learning task actually relates to the learning goals the teacher claims it addresses and the criteria used to judge student work.

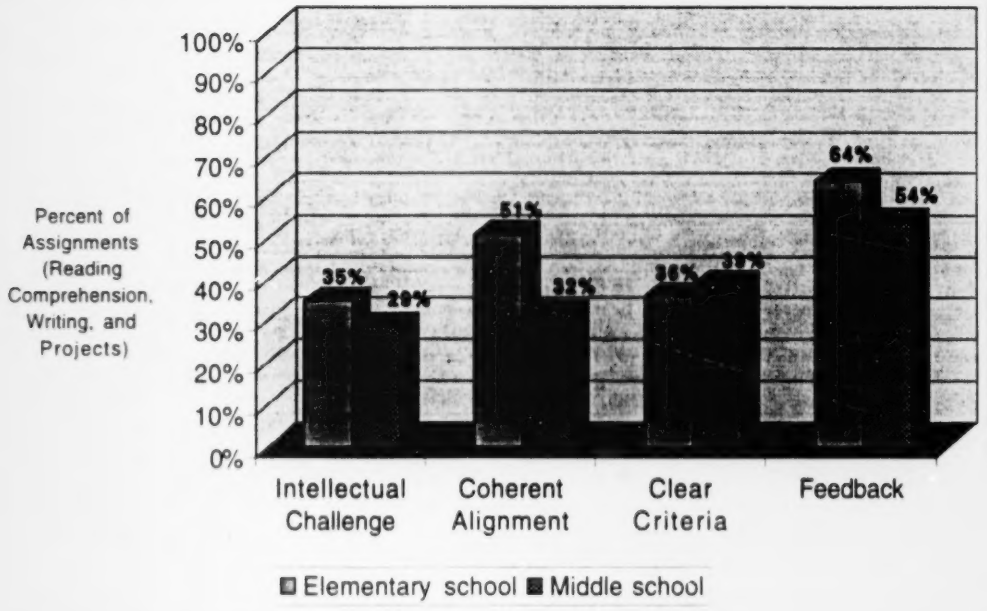


Figure 2. Learning environment profile.

success, and half to two thirds provided students with informative feedback. Overall, this profile suggests significant room for improvement in specific areas that might be addressed in future professional development. A brief discussion of these findings is described below to illustrate the value of the method in monitoring students' learning conditions.

Are students intellectually challenged? We used two variables to assess the extent to which students were intellectually challenged by the learning environment: (a) the "cognitive demands of the task"—whether an assignment required students to do more than make simple inferences or summaries (for example, by analyzing cause and effect, stating and defending opinions with facts, evaluating, or synthesizing information from several sources); and (b) the "knowledge required"—whether an assignment required students to use new or prior knowledge of some subject matter area or literature.

The vast majority of the assignments collected for this study at both elementary and middle schools made relatively low-level cognitive demands on students. Students were typically asked merely to recall information or use only moderately complex thinking like summarizing straightforward information, inferring a simple

main idea, or applying the appropriate writing format for a letter (i.e., such tasks were rated a 1 or 2 on the 4-point Cognitive Demands scale). If each type of assignment were to occur with equal frequency, then about 70% of elementary assignments and about 75% of middle school assignments did not ask students to think in very complex ways (beyond a rating of 2), as Table 9 illustrates. Higher cognitive demands (ratings of 3 or 4) occurred more than a quarter of the time in just three types of tasks: in 50% of the elementary reading comprehension tasks, in about 40% of the elementary content area writing tasks, and in 36% of the "challenging" projects at the middle school level.

Discipline-based content knowledge was a part of many elementary assignments (especially writing with a draft, content area writing, and challenging projects). However, middle school students were seldom expected to use discipline-based content knowledge in their assignments in English class, as illustrated in Table 10. Even with "challenging" projects, less than a quarter of the middle school assignments gave students the opportunity to learn to obtain and apply knowledge of any subject area. No doubt the different organization of schooling from elementary to secondary grades leads to this finding. Elementary teachers, responsible for the entire curriculum, often look for ways to relate the different subjects they must teach, such as having students practice reading comprehension, writing, and oral skills while learning social studies or science. Middle school English teachers, however, usually ask students to respond to literature rather than nonfiction.

Table 9

Cognitive Demands of Assignments

	Elementary % ≥ 3 rating (*n*)	Middle school % ≥ 3 rating (*n*)
Homework	22.7 (22)	26.0 (23)
Reading comprehension	50.0 (12)	25.0 (12)
Content area writing	41.7 (12)	NA[a]
Writing with a draft	16.7 (12)	18.2 (11)
Challenging project	22.2 (9)	36.4 (22)

[a] Content area writing assignments were not requested of seventh-grade English teachers since few of them teach interdisciplinary classes where such assignments might be found. Such assignments are far more common in elementary grades, where teachers often have students write about social studies or other content areas and grade their writing skills at the same time.

Table 10

Use of Discipline-Based Content Knowledge in Typical Assignments

	Elementary % of tasks calling for content knowledge (*n*)	Middle school % of tasks calling for content knowledge (*n*)
Homework	4.5 (22)	4.3 (23)
Reading comprehension	0 (12)	8.3 (12)
Content area writing	91.7 (12)	NA
Writing with a draft	41.7 (12)	9.1 (11)
"Challenging" project	66.7 (9)	22.7 (22)

At the middle school level, English class assignments more often called for students to read and react to fiction or poetry than to subject area knowledge, which was the reverse of the elementary school experience, as noted in Table 11. Over half the middle school writing-with-a-draft assignments called for students to read literature and write about it, whereas only a quarter of the elementary writing-with-a-draft assignments did. The same pattern was apparent in challenging projects, although the frequencies were much lower.

If intellectually challenging tasks are those with *both* higher cognitive demands (a rating above 2) and some requirement that students utilize some knowledge of literature or subject matter, then elementary tasks were more frequently "challenging" than middle school tasks. In addition, so-called "challenging major projects" were in fact more challenging than other types of middle school assignments, but this was not true at elementary schools. In the lower grades,

Table 11

Percent of Writing and Challenging Projects That Required Literary or Discipline Knowledge

	Elementary		Middle school	
Type of knowledge to be acquired or used in assignment	Writing with a draft (*n* = 12)	"Challenging" projects (*n* = 9)	Writing with a draft (*n* = 11)	"Challenging" projects (*n* = 22)
Literature	25	11	55	24
Discipline-based content knowledge	42	67	9	23
Format (of letter, essay, etc.)	67	33	55	51

reading comprehension and content area writing were more intellectually challenging by our definition than were the so-called "challenging major projects," as noted in Table 12.

Figures 3 and 4 are examples of middle school assignments that were rated high versus low on intellectual rigor. The "low" assignment requires very little, if any, knowledge of literature and the lowest level of cognitive demand because the assignment provided students with the answers to several questions during class discussion and cited page numbers, and the answers entailed merely one-line responses. The "high" assignment, on the other hand, required students to read multiple novels about a historical period, to synthesize and analyze substantive information from them, and to write articles in three different genres for a newspaper.

Table 12

Percent of Assignments Found to be Intellectually Challenging (High Cognitive Demands and Use of Knowledge)

	Elementary % of tasks (*n*)	Middle school % of tasks (*n*)
Homework	18 (22)	17 (23)
Reading comprehension	50 (12)	25 (12)
Content area writing	42 (12)	NA
Writing with a draft	25 (12)	18 (11)
"Challenging" project	30 (9)	45 (22)

No literature or content knowledge; cognitive demand = 1

- Answer 10 basic recall questions on a novel chapter read in class and use new vocabulary words from text. E.g.:
  - *What was Jamie's first decision as a treasurer? (p. 33)*
  - *What time did they reach the museum? (p. 36)*
  - *Find 3 words to describe Jamie's personality (p. 34, 35, 38)*
- (2 of the 10 questions were discussed in class and answers were put on the board; 3 more were discussed and answered orally before students wrote their answers)

Figure 3. Low-challenge middle school assignment.

Content knowledge (history) required; cognitive demand=4

- Use knowledge of WWII from reading several war novels and create a newspaper that includes 3 types of writing (from: cause and effect, biography, observation or evaluation), headlines, and an illustration.

Figure 4. High-challenge middle school assignment.

Are students given "coherent" assignments? The term "coherent" is used here to describe assignments in which the activity students do, the teacher's stated learning goals, and the criteria used by the teacher to evaluate student work are all aligned with each other. In other words, learning time is used on activities that should reasonably lead to the desired outcomes, and grading practices reinforce what is desired. It might seem odd that this notion is even addressed here, yet previous work (e.g., Aschbacher, 1994) suggests that this type of learning environment is more rare than one might expect. The results of the current study confirmed that students encountered "coherent" assignments less than half the time based on assignments submitted. Middle school students actually encountered "coherent" tasks less often (about one quarter to one third of the time) than did elementary students (about one fourth to over half the time), as illustrated by Table 13.

1. Tasks aligned with goals. In the vast majority of assignments in this study (75% to 80% of the tasks), what the students were asked to do was at least "partially aligned" with the teacher's stated learning goals (a 3 or better on our 4-point scale). The remaining 20% to 25% of the time, there was very little or no alignment, or the teachers' goals were so vague that alignment could not be determined (ratings of 1 or 2). For example, a teacher might say the goal of the assignment was to have

Table 13

Frequency of "Coherent" Assignments: With Aligned Goals, Tasks, and Criteria

	Elementary % of tasks (*n*)	Middle school % of tasks (*n*)
Homework	27 (22)	30 (23)
Reading comprehension	42 (12)	25 (12)
Content area writing	42 (12)	NA
Writing with a draft	50 (12)	36 (11)
Challenging project	60 (9)	36 (22)
Overall average	44	32

students learn to write an essay, but the task actually asked students to make an outline only. Elementary teachers tended to align their tasks and criteria to the learning goals better than middle school teachers did, and this was most pronounced with the challenging projects (for projects: elementary ratings averaged 3.72 on a 1-to-4 scale; middle school averaged 2.93; for writing assignments with a draft: elementary ratings averaged 3.54; middle school averaged 3.14).

2. Goals aligned with grading. Teachers' evaluation or grading criteria were not well aligned to their stated learning goals. In the best cases, about half of the elementary writing-with-a-draft and slightly over half of the elementary assignments challenging projects had criteria at least partially aligned with teachers' stated learning goals (a 3 or better on a 4-point scale). For all other task types at both grade levels, 60% to 75% of the assignments were rated as having little or no alignment between criteria and goals (ratings of 1 or 2). As noted for alignment of goals with learning activities, this rating was necessarily low when teachers did not have learning goals or criteria for performance in mind for a given activity or they could not articulate what they expected students to learn from a task.

How often did students encounter coherent, challenging assignments? Students were very seldom given assignments that were *both* "coherent" and "intellectually challenging," as illustrated in Table 14: about one assignment in six at elementary, one in ten in middle school.[8] Clearly, there is room for improvement in this instructional setting, and the methodology used here targets these areas for professional growth.

Table 14

Frequency of Assignments That Were Both "Coherent" and "Intellectually Challenging"

	Elementary % of tasks (*n*)	Middle school % of tasks (*n*)
Homework	9 (22)	0 (23)
Reading comprehension	25 (12)	8 (12)
Content area writing	17 (12)	NA
Writing with a draft	8 (12)	18 (11)
Challenging project	20 (9)	18 (22)
Overall average	16	11

[8] This frequency assumes that each type of assignment would occur with equal frequency in the real classroom, which is probably an overestimate of the frequency of higher quality assignments.

Are students given clear criteria for success? It seems reasonable to assume that students might apply themselves most effectively when they have a clear idea of what is expected, or what it takes to succeed at an assignment. Across all five task types and both grades studied, teachers tended to be rather unclear about their expectations for student performance (i.e., their grading criteria). Teachers tended to list a few dimensions such as "style, creativity, and punctuation" but left these terms completely undefined for students. The frequency of this vagueness varied across task types, as illustrated in Table 15, but such vagueness is probably most troublesome in assignments where students had to put in significant effort. In about three quarters of the "challenging" projects and over half to two thirds of the writing-with-a-draft assignments in this study, students were not provided clear guidance about how they would be graded. The greatest clarity among the tasks examined here was found in the elementary writing-with-a-draft assignments, where over 40% of the teachers described clearly, specifically and explicitly what they expected (received a rating of 3 or 4 on the 4-point scale used[9]). In only three tasks (all at middle school level), out of a total of 136, were students shown a model or concrete example of "good work."

Are students given informative feedback? Students received *no* feedback of any type in over a third of the assignments submitted at both elementary and middle school levels. Students were given no feedback on about half of the

Table 15

Percent of Tasks with Vague Expectations for Performance

	Elementary		Middle school	
	% Score of 1	% Score of 1.5-2.0	% Score of 1	% Score of 1.5-2.0
Content writing (E: n = 12; MS: n = 0)	18	55	—	—
Writing with a draft (E: n = 12; MS: n = 11)	8	50	18	46
"Challenging" project (E: n = 9; MS: n = 22)	11	67	23	55

Note. E = elementary; MS = middle school.

[9] Aschbacher, P. (September, 1998) *Looking carefully at classrooms.* Paper presented at the annual CRESST conference, Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

homework, half of the reading comprehension tasks, and two thirds of the writing in a content area (elementary). However, feedback of some type was given for the great majority of both the writing-with-a-draft assignment and the challenging projects (about 82% of the time overall). Feedback on these two task types took a variety of forms and varied from elementary to middle school, as illustrated in Table 16. The table shows the percent of assignments that provided each type of feedback. Some tasks provided more than one type of feedback.

Table 16 reveals a mixed picture of the learning environment in terms of the feedback available to help students learn. The good news is that teachers wrote comments or edited student work about half the time, although we made no attempt to evaluate the amount, quality, or usefulness of teachers' comments. The bad news is that, even after accounting for overlapping sources of feedback, students got feedback of questionable utility about one third of the time on these two types of assignments (i.e., a grade or unstructured peer edits with no other teacher comments or conferencing or rubric, or no feedback at all). Two potentially useful feedback strategies were very seldom used by teachers: structured peer feedback and rubrics, despite promotion of rubrics by the districts represented here. Several teachers commented in interviews that they did not yet feel comfortable creating or using rubrics.

How do teachers perceive student performance? We addressed this question by examining whether the student work that teachers submitted as examples of "high achievement" or "middle level achievement" was viewed similarly by raters

Table 16

Frequency of Feedback Given to Elementary and Middle School Students in Writing and "Challenging" Projects

	Elementary		Middle school	
	Writing-with-a-draft (n = 12)	"Challenging" project (n = 10)	Writing-with-a-draft (n = 11)	"Challenging" project (n = 22)
No feedback	8	30	18	18
Unstructured peer	67	20	18	27
Structured peer	17	0	9	9
Teacher notes/edits	67	40	45	50
Individual conference	33	10	0	5
Rubric score	17	20	18	18
Grade or points, no explanation	0	0	45	36

using a standards-based rubric developed for similar students (LAP rubric). Results of our analyses showed that correlations between teachers' views of student work and raters' LAP scores of the same work were low to moderate. As Table 17 illustrates, teachers' views were moderately correlated with the "Content" scale of the LAP rubric (with correlations of approximately .50) and only poorly correlated with the "Organization" and "MUGS" scales (approximately .25).

Table 18 shows the distribution of student work assigned "High" or "Moderate" labels by teachers arrayed alongside the scale of possible LAP total scores. In this analysis, it seemed reasonable to expect that "High" work could be expected to receive a total LAP score in the top third of the scale (a total score of 9 to 12 points—the top four possible LAP scores); "Moderate" work could be expected to receive a rating in the middle third of the scale (6 to 8 total points), and "Low" work, had we collected it, could be expected to receive a rating at the bottom third of the scale (3 to 5 total points). As the table indicates by the use of italicized letters, half (8 out of 16 essays) of the elementary work was rated one category higher by teachers than by the LAP ratings (i.e., a paper labeled "high" by a teacher received a LAP score in the "middle" range, 6–8). One third (9 out of 24 essays) of seventh-grade work was similarly judged higher by teachers than warranted by the standards-based language arts rubric used here. Thus teachers tended to view student work more favorably than did external raters. These results support the apparently low expectations for student work implied above by the relatively low levels of cognitive demands and use of content knowledge for most of the assignments.

Table 17

Correlations Between Teachers' Views and LAP Ratings of Student Work

Teachers' ratings using high/middle labels	LAP ratings		
	Content	Organization	MUGS
Elementary (n = 23)	.48	.18	.20
Middle school (n = 30)	.56	.32	.34
Overall teachers (n = 53)	.52	.24	.28

Note. LAP = UTLA Language Arts Project; MUGS = mechanics, usage, grammar, and spelling.

Table 18

Distribution of "Middle" and "High" -Rated Student Writing on Combined LAP Scale

	Possible LAP total scores (sum of three 4-point scales)	Distribution of student essays receiving given LAP score, each indicated by H for teacher's "High" rating or M for "Middle" rating[a]
Third grade		
LAP "High"	12	
	11	
	10	H H
	9	H
LAP "Middle"	8	*H H*
	7	*H* M M M M M
	6	*H* M
LAP "Low"	5	*H M M*
	4	
	3	
Seventh grade		
LAP "High"	12	
	11	H
	10	H H
	9	H H H H M
LAP "Middle"	8	*H H*
	7	*H H H* M M M M M
	6	M M
LAP "low"	5	*M M M*
	4	*M*
	3	

[a] Italicized letters indicate student work that teachers rated above or below its expected LAP score range.

Figure 5 displays similar information in a different format. It shows a similar pattern where teacher judgments were higher than rater judgments for both elementary and middle school writing. The graph illustrates that elementary teachers in this sample "overrated" student work more often than middle school teachers did.

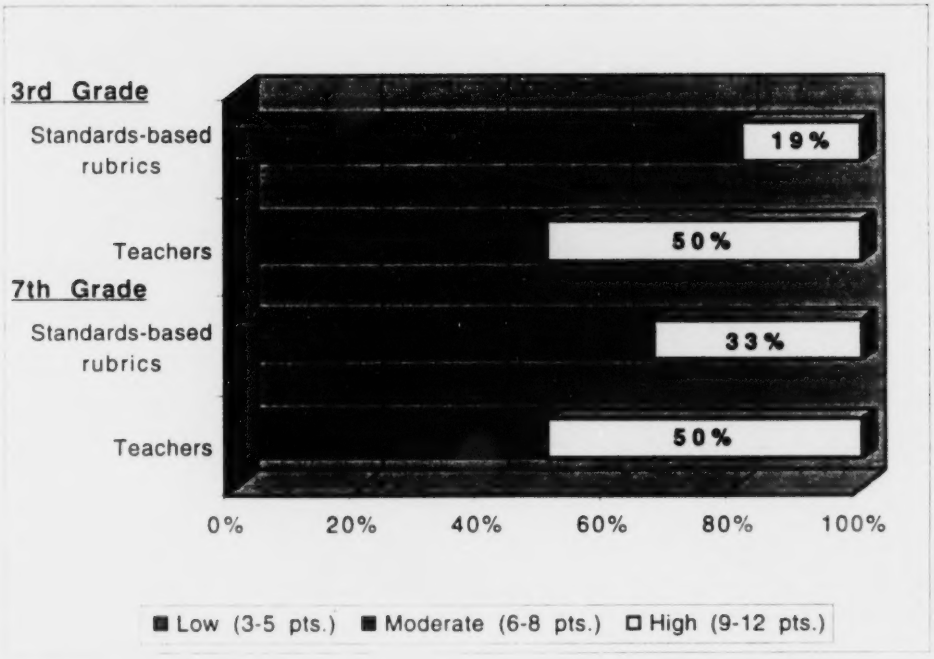


Figure 5. Teachers' views versus standards-based ratings of student work.

What can this methodology tell us about the relationship between the learning environment and student achievement? Challenge and achievement are correlated. We found that when teachers gave more challenging assignments (high cognitive demands and high overall quality), students performed at a higher level on writing assignments. We calculated correlations between students' LAP scores and ratings of their classroom assignments, and found small positive correlations (.34 to .43) between LAP scales and the rating on Overall Quality of Assignment, and between the Cognitive Demands scale for assignments and the Organization scale for LAP. All other coefficients were less than .30 (see Table 19). These findings suggest that "better" assignments and "better" student work have some tendency to occur together, but we cannot say whether either variable *leads* to the other. Both explanations seem likely to have some truth and are worthy of further research: (a) Teachers give more challenging assignments when they have stronger performing students in their class—and conversely, they give less demanding assignments when their students perform poorly; and/or (b) students do better when faced with

Table 19

Correlations Between Students' LAP Ratings and Ratings of Classroom Assignments

Assignment scales	LAP scales for student writing		
	Content	Organization	MUGS
Cognitive demands	.26	.38	.21
Clarity of grading expectations	-.06	.09	-.19
Alignment of goals to task	.26	.22	.29
Alignment of goals to grading	-.12	.25	.20
Overall assignment quality	.36	.34	.43

more challenging assignments. Certainly, students are almost sure to perform at a low level if they are not asked to use prior knowledge or acquire new knowledge for a task and to think in complex ways. When teachers give tasks that expect students to use their minds well, students at least have the opportunity to demonstrate their proficiency. Of course, merely providing such tasks does not ensure that students will do well. Good instruction is crucial.

Do experienced teachers give students better assignments? Are their assignments of more consistent quality? To investigate whether experienced teachers were more likely than inexperienced teachers to use assignments of consistent quality, we used a multiple regression analysis with the dependent measure being the standard deviation of the ratings summed across scales and assignments, with elementary and middle school teachers combined. We found that teacher experience in general did not predict consistency, but that the number of years teaching the specific grade level did, accounting for about 20% of the variance (adjusted R-square change = .20; $p < .05$). Given such a small sample size (24 teachers total), this result is interesting. It suggests that a teacher more familiar with a given grade level is slightly more likely to create assignments of a consistent quality level.

Did highest (or lowest) quality student work (based on lap ratings) occur in certain settings? We identified two middle school teachers and one elementary teacher whose students tended to have lower than average LAP scores and the same number of teachers whose students tended to have higher than average LAP scores. Then we examined the class characteristics of these teachers to see whether there were any distinguishing features. We found none. Class size, years of teaching experience, percent of students who had been in class since the beginning of the

year, percent of students with limited English proficiency, and average reading level of the class all were unrelated to student performance. We also quickly perused the assignment ratings for these teachers and found no apparent differences in such variables as cognitive demands, content knowledge, or clarity of grading expectations.

Did characteristics of the classroom relate to assignment quality? One potential use of an indicator of assignment quality would be to monitor the equity of educational settings. In this study we investigated possible relationships between the ratings of assignment quality and characteristics of the classroom such as class size, student stability (percent of students in the class since the beginning of the year or semester), proportion of the class who were limited English proficient, and the reading level of the class (according to teachers' self-report). We found no significant relationships, but this was influenced by the fact that there was little variability among the classes in the study. For example, only 2 of the 12 elementary classes had more than 20-21 students; only 3 had more than 20% of their students move during the year; and reading levels of all students in the 12 elementary classes were between 2.0–3.5. There was some variability in percent of students of limited proficiency in English (2 classes with about 25% LEP, 3 classes in the 50-70% range, and 7 at 100%), but this variable was not significantly related to assignment quality. Middle school classes had similarly low variability in size and stability. They varied somewhat more in percent LEP (0 to 100%) and in average reading level (2.0 to 7.5), but these were not significantly related to assignment quality. We cannot be sure whether the measure was insensitive or that classroom practice (as it occurred and was measured here) was not affected by the proportion of students with limited English proficiency or their average reading level. It remains to be demonstrated whether ratings of assignments might be influenced by class characteristics such as these.

Did teachers' reflections on assignments prove useful to them? It has become common over the past few years for teachers to come together to discuss student work. What has not happened so frequently, in our experience, is for teachers to have deep discussions to analyze their practice (e.g., goals for student learning, assignments, criteria, feedback strategies, and so forth) and to connect this practice to the student work it elicits. It was, in part, this concern that motivated this study. The interview data obtained underscore the importance of such conversations for teachers to improve practice. Several teachers explicitly mentioned that the

interview itself was a valuable forum for reflecting on their work, as illustrated by these comments from two teachers.

> It is only since you [the interviewer] have been asking me these questions that I'm also learning how to go back and reflect about everything I did wrong.

> I just wish that there was money available for small groups of teachers to get together and talk about their practice and bring samples of student work because that's so valuable. Just giving you these samples has really made me think about what I'm teaching. Too much of what we do is in isolation. I wish we could use the LAAMP money to break that isolation.

The open-ended interview experience allowed teachers to express their own frustrations and concerns about their practice and student performance—feelings they said they seldom shared in the typical school setting. The most common concerns were standards and rubrics. For example, one relatively new teacher seemed somewhat overwhelmed, and half of her students did not complete the writing assignment. During the interview she commented on her frustrations with standards and rubrics and then speculated on the connection to student performance.

> I have a hard time creating things around the standards. I like to give assignments and then look to see where it fits. This is because the standards are so general and open. If I did any assignment, I could find a standard that would align with it . . . It would help if the standards were more specific . . . We have a reporting information rubric I could have used. The rubrics are difficult for me to use . . . Next time I would let them [students] know the criteria ahead of time. I think I would probably get better results. They don't teach you much about assessment in your credential classes.

Another relatively inexperienced teacher commented during her interview about her problems trying to implement standards in the classroom despite district professional development.

> The real reason I did this [assignment] was so we could get something up on the board fast for open house. This is real typical of my teaching . . . The truth is I didn't prepare them [students] for the assignment . . . I don't use the standards because I don't know how to do it. I did almost no writing this year because I didn't know how to teach it. You hear all the time about the standards at the new teacher orientation. They pitch it all the time. Every time we take a class, the standards are brought up. They have us write down the standard we are using when putting together an assignment. It didn't guide me because it was too overwhelming. I don't even know how to teach much less put a standard to an operation that I don't know works. It's just one more thing I'm supposed

to do that I don't get . . . I started out without any thought. I just threw the lesson at them [students]. They started doing it, and I found myself really irritated at them for not doing what I wanted. Then I thought: What did I want? Well, I don't know what I wanted. They just weren't doing what I wanted them to do, and I didn't know what they were supposed to do.

A teacher of 8 years commented on the lack of support at her school for implementing standards in the classroom.

There is no mandate, no ongoing discussion at our school of applying standards. There was at the beginning of the year . . . Then there was no follow-up.

A teacher with over 20 years of experience reflected on a recent professional development experience related to standards and rubrics that had had a profound effect on her practice.

I look at the standards about twice a year. I looked at them more last year because I worked on a CRESST project that involved the standards. I don't do it on a regular basis now. They're written in a large and broad way. The times I do use them, I find them helpful because they help me keep track of how many students are meeting the standards . . . The CRESST/UTLA [LAP] rubric has had an extreme influence on me.[10] It was all about assessment guiding the curriculum. I learned that you need to keep your standards high and that students are not necessarily moving ahead just because they are going through the motions. It doesn't mean that students are reaching a standard . . . Hopefully teachers will get more specific about what you want to see, what you're looking for, and how do you get there. What happens when you don't get there. How to build it up so that you do get it, and make sure that children get enough practice over time so they approximate a standard.

Several teachers (of different levels of experience) made very interesting and potentially useful reflections on their practice during the course of the interview, such as the following:

That's another thing I might change next time: Copy some business letters and show them models.

I didn't tell them anything ahead of time [about criteria]. I would definitely let them know they need to include their web, and the criteria ahead of time. I think I would probably get better results.

---

[10] Actually, this teacher participated in a project that developed the LAUSD language arts standards and then developed curriculum, essay assessments, and the LAP rubric used in this study to score student writing. We suspect it was that experience, not just the rubric itself, that so influenced this teacher.

> The kids who read on a lower level tended to drop out . . . I think I needed more books on an easier reading level.
>
> It's a pain to have some accountability, but having you here also helps me to address what I'm doing, if I'm giving more feedback, less, reviewing the assignment, and looking at what students really learned and what I consider good or excellent work.
>
> I started out without any thought. I just threw the lesson at them. They had no background to do this . . . I had conferences and whole-class discussions. In the conferences I would tell them how to edit it, and they would come back with their papers unedited. It was just awful. I didn't teach it, and I didn't model it, so it was a mess. I just ended up doing it for them. The weakness in these papers has to do with the teacher . . . I would revise it for next year. I would really teach character development . . . I would break it out . . . and be really specific about character analysis.

Such introspection should not come merely when external researchers happen to visit the school to collect data on teaching. It highlights the value of guided inquiry and reflection for teachers on a regular basis.

Summary and Recommendations

Overall, our approach to measuring classroom practice through ratings of a sample of assignments shows promise in its capacity to describe several important aspects of the classroom learning environment. These findings appear useful in suggesting areas for administrative attention, professional development, and teacher reflection. Furthermore, teachers appreciated the opportunity to reflect on their practice through the questions we posed in our interview process, and they appeared to gain some insight into their teaching even though we did not structure the interviews for this purpose. Some of the interview questions used in this study might be incorporated into a school's self-evaluation process or teacher coaching based on the assignment rating scales. The technical quality of indicators based on this approach was adequate for this stage of the development process, but reliability of assignment ratings requires improvement in the future to reduce costs and improve generalizability. The next phase of this work will help determine which assignment types and rating dimensions are most useful in a very lean version for monitoring overall progress in school reform efforts in large-scale settings. Specific findings of the current study related to technical quality and utility are summarized below.

Technical Quality

The technical quality of this approach to classroom indicators was acceptable for such an early stage of development. To improve rater reliability, scale descriptions have been refined and anchor papers have been selected to illustrate most of the score points for each scale and will be used in rater training in the future. In addition, we recommend longer training of raters and the use of check papers to maintain rater stability and agreement throughout scoring.

Two of the six descriptive scales were particularly useful (knowledge applied in the task, and type of feedback provided) and should be used in rating the next set of data. All of the evaluative scales were useful, although two pairs of scales were significantly correlated ("Cognitive Demands" with "Overall Quality" and "Clarity of Grading Expectations" with "Alignment of Criteria With Tasks"). All of these scales should be retained for use with the next data set, and factor analyses and/or G-studies can be used to determine which scales are most helpful. G-studies conducted here revealed that in the next phase, a design using three to four assignments rated by two raters is important given the variability of these factors.

Teacher interview data generally supported the overall ratings of the assignments, although interviews clearly provided much more elaborate detail and enabled the interviewer to probe vagueness and apparent inconsistencies. Still, oral interviews, like written descriptions, can be vague or misleading. Future use of this approach should refine the written directions given to teachers and should consider increasing the incentives for them to complete assignment descriptions with detail and accuracy. Use of the technique by practitioners in the future for self-evaluation purposes is likely to be more engaging to them than mere participation in a low-stakes evaluation conducted by outside researchers.

Utility

The application of this methodology in eight schools as part of a program evaluation demonstrated that the approach enables us to describe the extent to which students encounter challenging, coherent assignments, high teacher expectations, clear criteria for achievement, and feedback to shape their learning. In this evaluation, for example, it revealed the following about the learning environment.

1. The vast majority of classroom assignments (tasks) from elementary and middle schools examined here were not intellectually challenging.

2. The majority of tasks were partially aligned with goals, but goals and criteria were frequently not aligned at all.

3. Only 1 in 6 elementary tasks and 1 in 10 middle school tasks were *both* intellectually challenging and "coherent" (i.e., tasks, goals, and criteria aligned).

4. In half the elementary students' writing and one third of the middle school students' writing, independent raters judged the work to be of lower quality than students' own teachers felt it was.[11]

5. Students were given unclear criteria for success in over half of the writing tasks and major projects.

6. Students saw models of what good work looks like in only 3 tasks out of 136.

7. Students received feedback on writing tasks and major projects, but it was of questionable utility since it seldom contained sufficient information to shape future learning.

8. Teachers seldom used grading rubrics despite their promotion by districts.

The findings based on the methodology used here make it possible to derive suggestions for professional development tied to specific problem areas, such as the following, for the schools examined:

1. how to raise teacher expectations for student achievement through familiarity with district rubrics and examples of excellent student work;

2. how to give students clear criteria for performance through rubrics or clear directions and examples; how to adapt rubrics to various assignments, and how to use them to evaluate their own students' work;

3. how to increase the intellectual challenge of assignments through the cognitive complexity of the activities students perform (cf. Marzano, 1992), and how to incorporate in some assignments the manipulation by students of content knowledge or literature (e.g., facts, ideas, concepts, principles);

4. how to increase alignment of student learning goals, activities, and criteria (and how to implement standards in the classroom);

5. how to give students useful feedback to shape learning; and how to guide students to provide structured feedback to peers that is accurate and

[11] Raters used a rubric, based on language arts standards for the students' grade level, that was developed by teachers and parents in the largest district studied.

helpful, while also helping them internalize rubrics by applying them to others' work.

In this study we also explored the method's capacity to identify possible relationships among characteristics of the classroom assignments, student work, and the classroom itself. We were able to show a couple of potentially interesting results. For example, there were some slight positive correlations between assignment ratings (overall quality and cognitive demands) and student work rated with the LAP rubric. Although we could not determine cause and effect, the fact that more challenging work was given to higher performing students poses an equity issue regardless of whether more challenging work is given to some students because teachers think they are more capable of it, or that students who are given challenging work are thereby encouraged to achieve at a higher level. This seems to be an area worthy of future research in which this methodology might be useful.

A second interesting finding was that although teaching experience was not related to teachers' consistency in assignment quality, experience at the particular grade level was. This result has implications for policy decisions about assignment of teaching staff. For example, it provides a concrete rationale for avoiding what was done at one school in the sample: assigning emergency credentialed teachers to all of the classrooms at a given grade level, leaving no colleagues to anchor new teachers' expectations of students at that grade. The value of grade-level experience could also be put to good use in formation of study groups and peer coaching situations. In a school where low expectations are entrenched, strategic assignments of staff to different grade levels could facilitate efforts to raise expectations.

We were unable to find significant relationships in these schools between classroom characteristics (such as class size and proportion of the class with limited English proficiency), assignment quality, and student achievement. Unfortunately, the lack of diversity among classrooms in this study and the small set of student work collected limited our capacity to explore such relationships with this data set.

One of our original goals was to develop a methodology for enhancing schools' and teachers' capacity to reflect on their practices and to improve themselves. At least half of the teachers interviewed in this study made unsolicited statements about the value of reflecting on their practice during the interview itself or demonstrated significant insights into their practice. Furthermore, they seemed to appreciate the opportunity to reflect, even though they were identifying areas of professional weakness. Since such spontaneous comments were not written on the

materials submitted to us, we conclude that the benefits of reflection were unique to the one-on-one interview setting. This result suggests the great potential value of including questions such as those in the interview protocol in study groups or other collegial professional development settings.

Next Steps

The results of this work suggest the following steps for future research.

1. Convert the "content knowledge" and "feedback" descriptive scales to evaluative scales (4-point scales).

2. Improve rater reliability in scoring student work in Spanish (through longer and more focused training, selection of more experienced raters, use of more examples); analyze results for all student work and for English and Spanish work separately.

3. Collect a larger sample of student work per classroom and explore the relationships among classroom assignment features, other aspects of the learning environment, and level of student performance.

4. Conduct G-studies to determine the most useful dimensions for rating assignments and the most useful assignments to collect and rate; conduct decision studies to identify the leanest design with sufficient reliability to determine the feasibility of using this approach in a large-scale setting.

5. Improve rater reliability in rating of assignments through refined scoring guides and additional anchor papers for each grade and each dimension (a sample of the anchor papers is appended to this paper).

6. Examine the standardized tests taken (e.g., SAT-9) and/or local standards and compare the content and cognitive processes in these documents to those called for in typical teacher assignments to see how well teachers are preparing students for the kinds of learning that are deemed important.

7. Continue to use this method to evaluate the quality of classroom assignments over time in selected LAAMP sites.

8. Have some practitioners pilot this method for self-evaluation; i.e., to collect, analyze, and reflect on a sample of their assignments and student work, such as in Critical Friends Groups; explore their perspective on the credibility and utility of this method.

9. Explore the extent to which ratings of a teacher's individual assignments are similar to or different from a more holistic rating of those assignments as a single body of work.

References

Aschbacher, P. R. (1992). *Humanitas Portfolio Project, 1991-92*. Los Angeles: University of California, Center for the Study of Evaluation.

Aschbacher, P. R. (1994). Helping educators to develop and use alternative assessments: Barriers and facilitators. *Educational Policy, 8*, 202-223.

Aschbacher, P. R. (1998). *Looking closely at classroom practice*. Presentation at the annual CRESST conference, University of California, Los Angeles.

Aschbacher, P. R., & Herman, J. (1991). *Final report of the Humanitas program evaluation, 1990-91*. Los Angeles: University of California, Center for the Study of Evaluation.

Aschbacher, P. R., & Rector, J. (1996). *The Los Angeles Learning Centers evaluation report. July 1994-June 1995*. Los Angeles: University of California, Center for the Study of Evaluation.

Baker, E. L. (1997). Model-based performance assessment. *Theory into practice, 36*, 247-254.

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice-Hall.

Baker, E. L., & Linn, R. (1998, Winter). Back to basics: Indicators as a system. *CRESSTline*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Blythe, R., Allen, D., & Powell, B. (in press). *Looking collaboratively at student work: A resource and guide*. Cambridge, MA: Harvard Project Zero and the Annenberg Institute for School Reform.

Carey, N. (1989). Instruction. In R. Shavelson, L. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education* (pp. 123-146; R-3742-NSF/RC). Santa Monica, CA: RAND.

Clare, L., & Valdés, R. (1999, April). The role of qualitative research in validating and informing indicator systems. In R. Ross (Chair), *Using indicator systems to evaluate large-scale school reform efforts*. Symposium conducted at annual meeting of the American Educational Research Association, Montreal, PQ, Canada.

David, J. L. (1988, March). The use of indicators by school districts: Aid or threat to improvement? *Phi Delta Kappan, 69*, 499-503.

Higuchi, C. (1996). *Improving student learning: High standards, standards-based curriculum, and standards-based assessment models.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Hoover, H. D. (1996, June). *The cost of performance assessments versus multiple-choice tests.* Paper presented at the annual CCSSO Large Scale Assessment Conference, Phoenix, AZ.

Los Angeles Compact on Evaluation. (1998, February). *Annual report: Los Angeles Annenberg Metropolitan Project.* Los Angeles: LACE.

Marzano, R. J. (1992). *A different kind of classroom: Teaching with dimensions of learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. A report to the nation and the Secretary of Education.* Washington, DC: U.S. Government Printing Office.

National Council on Education Standards and Testing. (1992). *Raising standards for American education. A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people.* Washington, DC: U.S. G.P.O.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report.* Chicago: Consortium on Chicago School Research.

Newmann, F. M., & Wehlage, G. G. (1995). *Successful school restructuring.* Madison: University of Wisconsin, Center on Organization and Restructuring of Schools.

Oakes, J. (1986). *Educational indicators: A guide for policymakers.* New Brunswick, NJ: Rutgers University, Center for Policy Research in Education.

Oakes, J., & Carey, N. (1989). Curriculum. In R. Shavelson, L. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education* (pp. 96-122; R-3742-NSF/RC). Santa Monica, CA: RAND.

Porter, A. (1988, March). Indicators: Objective data or political tool? *Phi Delta Kappan, 69,* 503-508.

Richards, C. E. (1988, March). Indicators and three types of educational monitoring systems: Implications for design. *Phi Delta Kappan, 69,* 495-499.

Selden, R. W. (1988, March). Missing data: A progress report form the states. *Phi Delta Kappan, 69,* 492-494.

Shavelson, R. J., McDonnell, L. M., & Oakes, J. (Eds.). (1989). *Indicators for monitoring mathematics and science education.* Santa Monica, CA: RAND/NSF.

Smith, M. S. (1988, March). Educational indicators. *Phi Delta Kappan, 69,* 487-491.

Appendix A

Student Work/Teacher Assignment Notebook (Elementary School)

1. Step-by-Step Instructions for Completing the Notebook
2. General Information Form
3. Assignment Cover Sheets: Reading Comprehension, Typical Writing Assignment with Final and Rough Drafts, Typical Content Area Writing Assignment, and Written Component of Very Challenging Assignment or Project

Directions for Collecting Assignments and Student Work
Step-By-Step Process: 3rd Grade Teachers

Due: May 8, 1998

Overview:

Please collect six assignments with four samples of student work for each assignment. You will be asked to fill out a cover sheet for each of the six assignments. The following gives you more detailed instructions.

1. **COLLECT THE FOLLOWING SIX ASSIGNMENTS BY MAY 8.**
   Between now and the end of April, collect **six** of the assignments you give third-grade students, with selected examples of student work. Use assignments which ask students to do some individual written work. Do not create new assignments specifically for this study. Please collect the following types of assignments:

   - **2** typical language arts homework assignments
   - **3** typical in-class assignments with a written response (one of each of the following):
     - 1 reading comprehension or reading response assignment
     - 1 writing assignment in a content area such as social studies, science, or math
     - 1 writing assignment that includes a rough draft and final draft, with any written feedback given by peers or teachers
   - **1** challenging major assignment/project with a written component

     If you have given or will give students a challenging major assignment or project that requires reading and has a written component, that is what we would like to see. You can use the most rigorous major assignment you gave or will give students anytime between January and May of this year. If this assignment has multiple steps, please submit only the written portion of the student work.

(continued)

2. **FOR EACH OF THE SIX ASSIGNMENTS COPY FOUR SAMPLES OF STUDENT WORK.**

   - **Choose two middle quality and two high quality pieces of student work from the same class.**
     It is fine to choose different students' papers for the different assignments. We just need two middle and two high for each assignment.
   - Copy the four pieces of student work for each assignment.
   - Place an ID sticker over each student's name. (We prefer to receive student work without their names so as to protect their privacy). Please do not cover up any part of the student's work, your feedback, or grade. If there is no clear area for the label, put it on the back of the work and cross out/white out the student's name.
     - Note: The student ID labels for Assignment #1 are stapled to the pocket for Assignment #1, and so forth.
   - Place an **M (Middle) or H (High)** sticker on each student paper accordingly. These stickers are in the plastic sleeve immediately preceding the blue pockets for student work.

3. **FILL OUT A COVER SHEET FOR EACH OF THE SIX ASSIGNMENTS.**

   Fill out the enclosed Cover Sheets for Teacher Assignments in the folders in this binder. There is a different cover sheet for each type of assignment, each on a different color of paper.

   - **Attach whatever will help us understand** the assignment and accompanying student work. (e.g., copy of the assignment given to students, rubric, outline of the unit, etc.).
   - Place the cover sheet with any attached papers and the four pieces of student work in the labeled folders at the back of this binder.

General Information Form
3rd Grade Teachers

Please answer the following questions.

1. How many years have you been teaching? _____ years

   a. How many years have you taught 3rd grade? _____ years

2. How many students are enrolled in your class? _____

3. Approximately what percentage of your students have been in your class since the beginning of the school year? _____ %

4. Please circle any of the following which describe your class:

   a. full bilingual b. modified bilingual c. SDAIE or sheltered English
   d. English only e. other _____

5. Approximately what percent of your students are LEP (Limited English Proficient)? _____ %

   a. In what language(s) do your LEP students receive language arts instruction? (Circle as many as apply.)

      English Spanish other _____

   b. Approximately what percent of your students have recently (within the past six months) been redesignated as Fluent English Proficient (RFEP)? _____ %

6. a. What is the range in reading level among your students? _____ grade to _____ grade

   b. At what grade level are <u>most</u> of your students currently reading? _____ grade

7. Is there anything else about your language arts class we should know when looking at the assignments and student work?

8. How similar is the language arts curriculum and instruction in your class to that of other teachers at your grade level in your school? (circle your answer)

not at all similar		somewhat similar		very similar
1	2	3	4	5

(continued)

9. What are the most important things you expect your students to be able to do by the end of the third grade in language arts? Please include what types of writing students are asked to do (e.g., narrative, descriptive, expository, persuasive, five-paragraph essays, etc.).

10. Has LAAMP influenced the kinds of assignments you give students, your level of expectations, or your grading practices? Please explain.

Thanks so much.

Date assigned: _____

Cover Sheet for Nightly Homework Assignment A

If you require more room to answer the questions, please use the back of this form.

1. **Describe the assignment below in detail or attach a copy of the assignment to this sheet.**

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help from a ❑ teacher ❑ teacher's aide ❑ other students ❑ parents (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it.
   If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

6. **Approximately what percent of students performed at the following levels on this assignment:**

   ____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Nightly Homework Assignment B

If you require more room to answer the questions, please use the back of this form.

1. **Describe the assignment below in detail or attach a copy of the assignment to this sheet.**

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help from a ❑ teacher ❑ teacher's aide ❑ other students ❑ parents
   (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it.
   If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

6. **Approximately what percent of students performed at the following levels on this assignment:**

   ____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Typical Class Reading Comprehension Assignment

If you require more room to answer the questions, please use the back of this form.

1. **Describe the assignment below in detail or attach a copy of the assignment to this sheet.**

   Specify the type (e.g., poem, novel, textbook, etc.) and grade level of the reading material. If students are working in reading groups, specify which group was given this assignment.

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)

   Students received help from a ❑ teacher ❑ teacher's aide ❑ other students ❑ parents (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

   Is this assignment an end-of-unit assessment? ❑ yes ❑ no

6. **Approximately what percent of students performed at the following levels on this assignment:**

   ____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Typical Class Writing Assignment: Final and Rough Drafts

If you require more room to answer the questions, please use the back of this form.

1. **Describe the assignment below in detail or attach a copy of the assignment to this sheet.**

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help from a ❑ teacher ❑ teacher's aide ❑ other students ❑ parents (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

Is this assignment an end-of-unit assessment? ❑ yes ❑ no

6. **Approximately what percent of students performed at the following levels on this assignment:**

____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Typical Class Content Area Writing Assignment

Please check one: ☐ science ☐ social studies ☐ math

If you require more room to answer the questions, please use the back of this form.

1. **Describe the assignment below in detail or attach a copy of the assignment to this sheet.** If students are reading as part of this assignment, please specify the level of the reading material.

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
Students received help from a ❑ teacher ❑ teacher's aide ❑ other students ❑ parents (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

Is this assignment an end-of-unit assessment? ❑ yes ❑ no

6. **Approximately what percent of students performed at the following levels on this assignment:**

____% = good - excellent ____% = adequate ____% = not yet adequate

Date assigned: _____

Cover Sheet for Challenging Major Assignment or Project: Written Component

If you require more room to answer the questions, please use the back of this form.

1. **Describe the overall assignment below in detail including the written component or attach a copy of the assignment to this sheet.** Specify the grade level of the reading material.

2. **What concepts, skills, and/or processes do you expect the students to acquire from this assignment?**

3. **How does the assignment fit in with your unit or what you are teaching in your language arts class this month?**

4. **What type of help, if any, did students receive to complete the assignment?** (Check all that apply.)
   Students received help from a ❑ teacher ❑teacher's aide ❑other students ❑parents (e.g., substantive revision feedback from teacher or peers). Please explain:

5. **How is this assignment assessed?** If there is a rubric, student reflection, etc., please attach it. If you are not attaching a rubric, please explain your criteria for deciding which papers are middle papers and which are high.

Is this assignment an end-of-unit assessment? ❑ yes ❑ no

6. **Approximately what percent of students performed at the following levels on this assignment:**

   ____% = good - excellent ____% = adequate ____% = not yet adequate

Appendix B

Rubrics for Scoring Teachers' Language Arts Assignments

1. Version 1 for Spring 1998
2. Version 2 for Spring 1999

RUBRIC FOR SCORING TEACHERS' LANGUAGE ARTS ASSIGNMENTS v. 1[12]

Language Arts Assignments—Descriptive Categories

TYPE OF ASSIGNMENT	TYPE OF CONTENT KNOWLEDGE	RESPONSE ELICITED	STUDENT CHOICE	SCORING DIMENSIONS	TYPES OF FEEDBACK
1. reading comprehension 2. expository writing 3. persuasive writing 4. creative writing/stories 5. journal writing 6. grammar 7. spelling 8. vocabulary 9. other	1. knowledge of format/genre/poetry or literary vocabulary 2. personal knowledge/reflection 3. fiction/poetry 4. discipline-based knowledge 5. research or study skills, strategies 6. vocabulary 7. none	1. true/false 2 multiple choice 3. one-word/fill-in-blank 4. short response (a few words or couple of sentences) 5. paragraph 6. extended response/essay/ multiple paragraphs	1. no choices in task 2. restricted choice in topic (e.g., choose from one of three topics) 3. open ended choice: e.g., create own essay topic 4. choose resources to use (book to read, library, internet, etc.) 5. other choice	1. mechanics, usage, grammar, spelling, sentence structure 2. organization 3. use of specified resources; details; thoughtfulness, inclusion of specified information 4. application of format or style for genre 5. creativity/originality 6. turned in or not 7. other 8. NONE mentioned in cover sheet, rubric, directions	1. no feedback or no info to tell 2. unstructured peer feedback 3. structured peer feedback 4. written comments, edits 5. individual conference 6. small group conference 7. whole class discussion 8. rubric score 9. letter grade or points

[12] Version of rubric used in spring 1998.

Language Arts Assignments Rubric v.1

COGNITIVE DEMANDS **(cover sheet & student work)**	**4** Task requires strongly complex thinking as an extensive, major focus of task. Student may be asked to analyze cause and effect, identify a problem and pose reasonable solutions, speculate giving details or justification, defend opinions or argue a position with evidence, etc. to a great extent	**3** Task requires some strongly complex thinking. Student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate giving details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or relevant from unimportant or irrelevant); determine bias, values, intent.	**2** Task requires moderately complex thinking. Student may be asked to summarize straightforward information, infer simple main idea, apply the appropriate format for a given genre.	**1** Task requires only recall of basic information. Student may be asked to answer simple reading comprehension questions, remember facts, recall definitions. Writing on a topic with little focus or structure; no reading required.
GRADING EXPECTATIONS **(cover sheet & rubric)**	**4** Teacher's grading criteria are clear and explicit, for example, a model of good work is shown to students. The sample of ideal or previous students work may be provided during instruction (immediately) preceding the task.	**3** Teacher describes clearly, specifically, and explicitly what is expected with little or no question. Teacher may use a rubric or other guidelines such as definitions or examples, but not simply a list of features.	**2** Teacher provides some general directions or a rudimentary rubric. E.g., a list of dimensions such as "style, creativity, and organization", but some or all dimensions are undefined or vague.	**1** Not clear from teacher's documents what the teacher expected. Teacher may even say that s/he did not think about grading at the time of assigning the task.
MATCH BETWEEN LEARNING GOALS AND TASK **(cover sheet & task sheets)**	**4** There is exact alignment between (a) teacher's stated learning goals for students on that assignment and (b) what the task asks students to actually do E.g., goal is being able to summarize several points and activity entails summarizing; tasks and goals overlap completely—neither one calls for something not included in the other.	**3** There is some or partial alignment between teacher's stated learning goals and what the task asks students to do. E.g., goal is to recall details and summarize information but task only calls for recall of details, or includes MUGS not mentioned in goal.	**2** There is very little or no alignment between teacher's stated goals and what the task calls for, but there may be some minimal relationship between them. E.g., goal is to be able to write an essay, but task calls for completing a concept map and making an outline for an essay (but NOT actually writing an essay).	**1** There is no alignment or relationship between teacher's stated goals and what the task calls for students to do. —or teacher does not state any goals —or goals are so vague that alignment cannot be determined —or teacher's goal isn't what s/he claims it is, e.g., "writing to an outline." E.g., goal calls for writing an essay, but task calls for giving an oral report.

Language Arts Assignments Rubric v. 1 (con't)

MATCH BETWEEN LEARNING GOALS AND STATED GRADING **(cover sheet & rubric)** **(do not bother to look at student work)**	**4** There is exact alignment between (a) teacher's stated learning goals for students on that assignment and (b) teacher's stated grading criteria. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions such as stating a point of view and providing relevant supporting evidence; do not include dimensions not mentions in goals (e.g., creativity). (MUGS may be considered relevant when writing for authentic audience even if omitted from goals.)	**3** There is some or partial alignment between teacher's stated learning goals and the stated criteria for grading. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions but also extraneous ones. Or, fails to include critical dimension (e.g., support for assertions or point of view).	**2** There is little alignment between teacher's stated learning goals and the stated grading criteria. E.g., goal is to write a business letter, but criteria include <u>mostly</u> extraneous dimensions, e.g., participation in class discussion is given more weight than letter format. Or, criteria given are not very appropriate, e.g., slang is acceptable in a business letter.	**1** There is no alignment between teacher's stated learning goals and the stated grading criteria. Or, there may be no stated goals, or no stated criteria (thus no alignment possible).
OVERALL TASK QUALITY **take into account:** - **if homework assignment** - **time of year assigned** - **purpose of task** - **quality of student response elicited**	**4** Excellent quality in terms of appropriateness/worthiness of goal, application of goal in task, and scoring criteria. I.e., High-level cognitive processes are elicited, reading is rigorous, content is analyzed at a deep level. Students are asked to use their minds well in a task that is likely meaningful to them and of long-term use in building useful knowledge and skills.	**3** Good quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Higher level cognitive processes are elicited, good quality literature is read, discussion of ideas is higher-level.	**2** Limited quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Low-level cognitive skills and/or content knowledge is required; poor materials/resources may be used; pedantic approach to teaching.	**1** Poor quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Task may require prior knowledge not taught by teacher or that should not be expected of students; Students are not expected to use their minds well. Often a task with very minimal expectations, or with very limited value.

RUBRIC FOR SCORING TEACHERS' LANGUAGE ARTS ASSIGNMENTS (v. 2)[13]

Language Arts Assignments—Descriptive Categories

TYPE OF ASSIGNMENT	RESPONSE ELICITED	STUDENT CHOICE	SCORING DIMENSIONS	TYPES OF FEEDBACK
1. answering set of reading comprehension questions 2. expository writing 3. persuasive writing 4. creative writing/stories 5. journal writing 6. grammar 7. spelling 8. vocabulary 9. other	1. true/false 2. multiple choice 3. one-word/fill-in-blank 4. short response (a few words or couple of sentences) 5. paragraph 6. extended response/essay/ multiple paragraphs	1. no choices in task 2. restricted choice in topic (e.g., choose from one of three topics) 3. open-ended choice, e.g., create own essay topic 4. choose resources to use (book to read, library, Internet, etc.) 5. other choice	1. mechanics, usage, grammar, spelling, sentence structure 2. organization 3. use of specified resources, details, thoughtfulness, inclusion of specified information 4. application of format or style for genre 5. creativity/originality 6. turned in or not 7. other 8. NONE mentioned in #5, rubric, directions	1. no feedback or no info to tell 2. unstructured peer feedback 3. structured peer feedback 4. written comments, edits 5. individual conference 6. small group conference 7. whole class discussion 8. rubric score 9. letter grade or points

[13] Revised rubric for use in spring 1999.

Language Arts Assignment Rubric v. 2

COGNITIVE DEMANDS **(cover sheet & student work)**	**4** Task requires strongly complex thinking as an extensive, major focus of task. Student may be asked to analyze cause and effect, identify a problem and pose reasonable solutions, speculate with details or justification, defend opinions or argue a position with evidence, etc. to a great extent.	**3** Task requires some strongly complex thinking. Student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate with details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or relevant from unimportant or irrelevant); determine bias, values, intent.	**2** Task requires moderately complex thinking. Student may be asked to summarize straightforward information, infer simple main idea, apply the appropriate format for a given genre.	**1** Task requires only recall of basic information. Student may be asked to answer simple reading comprehension questions, remember facts, recall definitions. Writing on a topic with little focus or structure; no reading required.
CLARITY OF THE TEACHERS' GOALS **(cover sheet and assignment directions)**	**4** Goals are all very clear and explicit in terms of what students are to learn from the assignment. Additionally, all the goals are elaborated, and framed in terms of student learning.	**3** Goals are mostly clear and explicit in terms of what students are to learn from the assignment, and are mostly framed in terms of student learning. Some activities may be included.	**2** Goals are somewhat clear and explicit in terms of what students are to learn from the assignment. Goals may be a combination of goals and activities with no definable objective.	**1** Goals are not clear in terms of what students are to learn from the assignment OR all goals may be stated as activities with no definable objective ("activity for activity's sake").
CLARITY OF GRADING EXPECTATIONS **(cover sheet & rubric)** **Note: Only score if we have teachers' grading criteria.**	**4** Teacher's grading criteria are clear and explicit (for example, a model of good work is shown to students). The sample of ideal or previous student work may be provided during instruction (immediately) preceding the task.	**3** Teacher describes clearly, specifically, and explicitly what is expected with little or no question. Teacher may use a rubric or other guidelines such as definitions or examples, but not simply a list of features.	**2** Teacher provides some general directions or a rudimentary rubric. E.g., a list of dimensions such as "style, creativity, and organization," but some or all dimensions are undefined or vague.	**1** Not clear from teacher's documents what the teacher expected. Teacher may even say that s/he did not think about grading at the time of assigning the task.

Language Arts Assignments Rubric v. 2 (con't)

COHERENCE OR MATCH BETWEEN LEARNING GOALS AND TASK **(cover sheet & task sheets)** **Note: Only score if we have teachers' goals.**	**4** There is exact alignment between (a) teacher's stated learning goals for students on that assignment and (b) what the <u>task</u> asks students to actually do, AND task fully supports instructional goals. E.g., goal is being able to summarize several points and activity entails summarizing; tasks and goals overlap completely—neither one calls for something not included in the other.	**3** There is some or partial alignment between teacher's stated learning goals and what the task asks students to do, AND task supports instructional goals. E.g., goal is to recall details and summarize information but task only calls for recall of details, or includes MUGS not mentioned in goal.	**2** There is very little alignment between teacher's stated goals and what the task calls for, AND task somewhat supports or furthers instructional goals. E.g., goal is to be able to write an essay, but task calls for completing a concept map and making an outline for an essay (but NOT actually writing an essay).	**1** There is no alignment between teacher's stated goals and what the task calls for students to do (e.g., goal calls for writing an essay, but task calls for giving an oral report), AND task does not support or further instructional goals. —OR goals are so vague that alignment cannot be determined.
MATCH BETWEEN LEARNING GOALS AND STATED GRADING **(cover sheet & rubric—not student work)** **Note: Only score if we have teachers' goals <u>and</u> grading criteria.**	**4** There is exact alignment between (a) teacher's stated learning goals for students on that assignment and (b) teacher's stated grading criteria. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions such as stating a point of view and providing relevant supporting evidence; do not include dimensions not mentioned in goals (e.g., creativity). (MUGS may be considered relevant when writing for authentic audience even if omitted from goals.)	**3** There is some or partial alignment between teacher's stated learning goals and the stated criteria for grading. E.g., goal is to write a persuasive essay, and criteria include appropriate dimensions but also extraneous ones. Or, fails to include critical dimension (e.g., support for assertions or point of view).	**2** There is little alignment between teacher's stated learning goals and the stated grading criteria. E.g., goal is to write a business letter, but criteria include <u>mostly</u> extraneous dimensions, e.g., participation in class discussion is given more weight than letter format. Or, criteria given are not very appropriate, e.g., slang is acceptable in a business letter.	**1** There is no alignment between teacher's stated learning goals and the stated grading criteria.
OVERALL TASK QUALITY **Take into account all previous dimensions plus:** **- time of year assigned** **- purpose of task** **- quality of student response elicited**	**4** Excellent quality in terms of appropriateness/worthiness of goal, application of goal in task, and scoring criteria. I.e., High-level cognitive processes are elicited, reading is rigorous, content is analyzed at a deep level. Students are asked to use their minds well in a task that is likely meaningful to them and of long-term use in building useful knowledge and skills.	**3** Good quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Higher level cognitive processes are elicited, good quality literature is read, discussion of ideas is higher level.	**2** Limited quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Low-level cognitive skills and/or content knowledge is required; poor materials/resources may be used; pedantic approach to teaching.	**1** Poor quality in terms of appropriateness of goal, application of goal in task, and scoring criteria. I.e., Task may require prior knowledge not taught by teacher or that should not be expected of students. Students are not expected to use their minds well. Often a task with very minimal expectations, or with very limited value.

Appendix C

Teacher Interview Protocol
Class Assignment/Student Work Interview Questions

Class Assignment/Student Work Interview Questions

Teacher ID: _____ Interviewer: _____

School ID: _____ Grade: _____ Subject: _____

1. Did you create this assignment yourself? If not, where did it come from? (e.g., select it from a textbook, get it from a colleague, jointly plan with colleagues, other?)

   Have you ever used it (or a version of it) before?

   If yes: With a class like this one or in a different grade?
   Different types of students?
   What kinds of changes have you made since the first time you used it?

   How often do you usually give assignments like this one to your class?

   Why did you create or select this assignment? What appealed to you about it?

2. Tell me about the instructional context for this assignment—what you taught leading up to and immediately following this assignment, i.e., describe how it fits into your overall class.

   Was this a culminating activity of a particular unit of instruction?

3. What did you want your students to learn or be able to do from this assignment? (i.e., learning goals for students—cognitive, affective, metacognitive, social learning, etc.)

4. Did students use any technology in this assignment? (video, computers, etc.) (Note: Do not imply they should have used technology in the assignment.)

   If yes: How did you want students to use technology in this assignment?
   Why did you incorporate technology into this assignment? (e.g., part of standard, mandate, their own idea, new technology from LAAMP money, etc.)

5. (Check prior to interview what teacher and/or principal said in any previous interviews about standards that may be emphasized by school, Family, or district.) Did you have any standards (school, district, state, national, other) in mind when you planned this assignment?

   If yes: To which actual standard(s) does this particular assignment relate? (i.e., have teacher recite or literally show you one or more standards that relate to this assignment so you can see if they seem aligned)
   How often do you usually plan your assignments and learning activities around these standards?
   Why do you use these standards? (e.g., school mandate, School Family decision, district mandate, teacher's own idea, other)

   If no: How did you proceed? Why weren't the standards helpful? (e.g., personal decision; no mandate or encouragement from school, School Family, district; standards too vague to be helpful; etc.)

6. How did this assignment work out?

Did most of your students seem engaged in it? (i.e., try hard or enjoy it)

How long did it take most students to work on it? (List in-class time and out-of-class time.)

How well did students do on it? (Record % or fraction of class below.)

At what grade level are each of these three groups working? (Record below).

	Proportion	Grade level working at
excellent or good		
adequate		
really poor or failing job on it		

What problems, if any, did students tend to have with it and why?

At which grade level was the activity aimed?

7. How did you grade or evaluate students' work?

What criteria or rubric, if any, did you use?

Where did your criteria (rubric or grading guidelines) come from? (e.g., self-created, jointly with colleagues? LAAMP, school, or district rubrics? students help determine criteria?)

What did you tell students about how you would grade or evaluate their work?

Did you show them any examples of what "good" work looks like?

Let's look at the samples of average and excellent student work on this assignment. Are these pretty typical?

What makes these two papers "good" work? (e.g., how can you tell they "get it"?)

What makes these two papers "average" work?

What kinds of mistakes or problems did students who performed poorly make?

What other things did you take into account in grading individual students? (e.g., personal growth over time, effort, behavior, participation, compared to specific objectives, compared to others in class)

8. After you saw what your students did on this assignment, did you use that information in any particular way for yourself or your students? (e.g., to change what you teach next, revise the assignment for next time, plan remediation for certain students, etc.)

Can you give a specific example for this assignment?

Would you make any changes on it next time? (or possibly not use again?) If so, please describe changes you would make.

9. What kind of feedback did you give students on this assignment other than the grade, if any? (e.g., written and/or oral comments)

   Did they get any feedback before the final draft? (i.e., from peers or you? get a chance to revise?)

   What types of comments do you typically make?

10. Did LAAMP professional development or other program elements influence your selection or use of this assignment in any way? (e.g., kind of assignment, level of expectation, standards alignment, grading rubrics or practices, joint planning with colleagues, etc.)

11. If you were asked to help a group of new teachers at your school create some good assignments for their classes:

    What would you tell them are some key features of good assignments?

    What should these new teachers know about how students learn that will help them create good assignments?

12. Has your approach to language arts been changing? (i.e., any different ideas about what literacy is and how to help students achieve it)

    If so, how is your approach changing?
    Why? (Is any change related to LAAMP or recent professional development?)

13. Is there anything else you'd like us to know about your class, about teaching in this school, or about LAAMP?

Appendix D

Class Assignment Anchor Paper Summaries

Cognitive Demands
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	100504 400	This science task was scored a "1" for cognitive demands. Students were asked to answer a prompt about a science project they had completed at home. The prompt asked students to do the following: Write about how they did the project, where they got the materials, who helped them, and whether it was difficult or easy to do. Students were given a set of directions to follow at home to implement the experiment. After carrying out the experiment, students answered basic recall questions that were not connected with science content or science thinking. Students were not required to explain, infer, generalize, or synthesize. The questions did not require any complex thinking. Because of this, the demands of the task were low level. This task was actually given to second graders, but we are using it here as an example of an assignment that would have been a "1" for cognitive demands had it been given to third graders.
2	010302 500	In contrast to the previous scored assignment, this third grade science task required some moderately complex thinking and was scored a "2." Students were asked to observe Brine Shrimp and take notes answering specific questions: (1) Can you see legs? (2) Do they stay together? (3) Do they float? (4) Do they ever bump into each other? (5) Do they like one side of the container more than another? (6) If we leave the unhatched eggs in, do you think they will hatch? (7) Do you think there is something else growing in your cup? And (8) Can you tell the difference between a male or female Brine Shrimp? Then students had to use their notes to write a report about their Brine Shrimp. This task only required some moderately complex thinking. Questions six and seven required students to make inferences. The same task could be constructed differently, however, to require more strongly complex thinking. For example, the task could ask, "Based on your observation, how to you think Brine Shrimp move from place to place," or "Do you see evidence of how Brine Shrimp reproduce?" This kind of question would force students to make an observation and connect it to a function and would raise the cognitive demand score. In this task, the student work revealed that students tended to answer the questions in yes and no fashion, without supporting their answers. Students were not required to explain their thinking or tie their thoughts to a rationale. The task did not require students to make complex connections.
3	010301 300	In this reading comprehension assignment, third-grade students were asked to predict how a party given by Beezus in the book Beezus and Ramona would be similar or different from the party given by Ramona in the book. This assignment was scored a "3" for level of cognitive demand since it required that students analyze and compare characters from the book. Students were required to understand the book, speculate or make a prediction, and use details from the text as evidence to justify that prediction.

Score Point	Assignment ID	Assignment Description and Comments
4	160501 600	This challenging major assignment required students to create a science experiment and write up each step of the scientific method. This assignment was scored a "4" since it required strongly complex thinking such as designing an experiment, forming a hypothesis by identifying a problem and posing reasonable solutions, and analyzing cause and effect in the discussion of the results. The task involved communicating all of this in a coherent way according to a format. A student's analysis might include some speculation such as when a student comes to the conclusion that more testing needs to be done or that the results require a new hypothesis. This assignment illustrates that even assignments that call for strongly complex thinking can result in lower level thinking if students simply put little effort into it, or if the setting is not sufficiently structured to guide students to this level of thinking. Because students completed this project at home, many students did not have an expert available to help guide their learning. For example, one student did an experiment to see if the heavy liquid would go to the bottom and the lighter liquid would go to the top. He did not do any measurements of volume or mass to determine which was "lighter" or "heavier" because he had no way of knowing that he had chosen an experiment on density. The student made an assumption that because the syrup was at the bottom, it was the "heaviest." There should have been an expert to either suggest a change in his experimental design (e.g., that he take a fixed volume of each liquid and weigh it) or to steer the student toward a completely different area more suited for his level and one that he could learn from. Because the student didn't do these measurements at the outset, his analysis was circular and constrained, and the amount of knowledge he got from the experiment was quite limited. The student would have gotten a clearer picture of the scientific method if he had first conferred, because that is part of what scientists do.

Clarity of Teacher's Goals
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	160502 100	This is an example of a 1 in terms of Clarity of Teacher's Goals. For this assignment, students were given a list of vocabulary words and asked to write one sentence about each one. The teacher's stated goals were as follows: *Good sentence structure. I can assess their needs and understanding of word meaning.* These goals are difficult to interpret in terms of what students are supposed to learn. If the goals are to have students learn or practice good sentence structure or learn the meaning of particular words, it is not clear what benefit students derive from these activities.
2	160501 300	The following goals are an example of a 2 on the Clarity of Teacher's Goals scale. The assignment was for students to write a retelling of a story after reading the story from a basal reader or listening to the story being read to them. The teacher stated the goals for this assignment as follows: *I want them to acquire comprehension skills and develop in writing skills as well. However, most important is how much information they comprehend* These two goals—acquiring comprehension skills and developing writing skills—are stated in a relatively clear but general way. That is, we know what the general skills are that the teacher is targeting, but we do not have sufficient detail to understand how writing a retelling of a story will improve a students' reading comprehension nor how it will help a students' writing. The goals are stated too broadly.
3	160501 600	The following is an example of a 3 in terms of Clarity of Teacher's Goals. For the assignment, students were to conduct a science experiment adhering to the scientific method and then write up their experiment. The teacher wrote that the goals of this assignment were the following two: *To display their understanding of each step of the scientific method and learn to enjoy experimenting.* This pair of goals is generally clear but not elaborated to the point that the reader knows beyond a doubt what the teacher had in mind for student learning. Both goals are slightly general, and some elaboration, perhaps specifying the steps of the scientific method or stating some benefits of learning to enjoy experiments, would have been helpful. Also, the teacher does not mention writing goals specifically, and mention of how students' writing would be improved would also have been useful.

Score Point	Assignment ID	Assignment Description and Comments
4	080901 600	The goals below are an example of a 4 in terms of Clarity of the Teacher's Goals. The assignment asked students to write a five-page report on an animal of their choice by answering teacher-generated questions such as "How are its young born? Are they hatched from eggs or are they born alive from their mothers' bodies?" In completing the assignment, a student wrote a rough draft, revised it with help from a peer, submitted a revised copy to the teacher for editing, and wrote a final draft. The assignment and its responses were written in Spanish. The teacher's stated goals were as follows: *How to write a report. Practice revision. Develop the skill of using reference material to find answers. Learn the concepts of life cycle, food chain (whether their animal is prey, predator, or both), habitat, and the idea that different animals live on different continents.* Each of these four goals is clear and specific about what students are to gain from the assignment. The goal having to do with the life science content that students are supposed to learn is particularly explicit. Though the goals are not elaborated, the way in which they are stated provides sufficient information as to leave no question about their meaning. In addition to being clear and specific, the individual goals form a coherent set of goals that together work toward teaching students the skill of writing reports.

Grading Expectations
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	160501 400	For this third grade science task, students were asked to observe and record the movement of three different objects, one light, one medium, and one heavy. They also observed and recorded the motion of a wind-up car. This task is part of a science unit on force given to the teacher. The lesson plan states that part of the assessment is students' understanding of what the words push and pull mean. This task was scored a 1 for clarity of grading expectations partly because the teacher had not thought about the grading when the assignment was given. In addition, the only criterion used after the fact for scoring student responses was whether students used the words "push" or "pull" in their responses. The teacher gave higher scores to students who used the words "pushed" and "pulled," regardless of whether those responses showed an understanding of the scientific concept of force. One student who received a high score wrote: "The pencil moved because Julien pushed it down. The chair moved because Alex pulled it. The shelf moved because Caesar pushed and pulled it." The grading expectations were not clear and were not made explicit to students.
2	160502 300	This third grade reading comprehension task was scored a 2 for clarity of grading expectations. This task required students to write a retelling of the story My Great Aunt Arizona after listening to it read aloud. The teacher's grading criteria are very general: (1) Did they understand what was read? (2) Can they write a beginning, middle and end? These dimensions are undefined. What does a student need to do to show that he/she understands what was read? How much detail does the teacher expect students to include. Is the retelling a straight recall or does the teacher expect the student to add something to show deeper understanding of the story's meaning. There is so much left open, that unless the teacher modeled with lots of examples of good work, students might not know what exactly they need to do to fulfill the task requirements.
3	990502 500	In this third grade challenging major assignment, students were asked to select a famous person to research and take notes on while reading in the following seven areas: birth and death, childhood, education, older years, why famous, other interesting information. This task was scored a 3 for clarity of grading expectations because the teacher describes clearly, specifically, and explicitly what is expected. The criterion for a high score was: Each of the seven sections contains accurate information and has several details above and beyond the minimum. The student uses correct sentence and paragraph structure In the original notes, the student did an excellent job summarizing the important information about the famous person. The criterion for a middle paper were: Most of the information is accurate. It may include minor errors. There is some information in each of the seven sections. There are at least three details for each section. It is written in complete sentences. The student at least attempts paragraph structure.
4	N/A	N/A

Match Between Learning Goals and Task
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	N/A	N/A
2	160502 400	This task was scored a "2" for coherence/match between learning goals and task. The teacher's goals were "writing skills, recall, develop memory, and remember important facts." Following a lesson on volcanoes, the teacher asked students to write down what they knew about volcanoes The teacher stated that many of the facts students were asked to recall were actually written on the board while students were writing, negating one of the main goals for giving the assignment. If not for this, the task would have been scored a "3." Also, the teacher listed "writing skills" as a goal, which is general and vague. The teacher does not clearly state how students will develop writing skills from the assignment, unless what she wants is for students to simply have practice writing in complete sentences. We can't know from what is written. This task would have been scored a "4" if many of the facts had not been left on the board and if the teacher had stated the writing skills goal with more clarity and specificity.
3	N/A	N/A
4	010302 100	This homework assignment required students to do research and write a two-page history report on a famous woman. The teacher's goals were "to strengthen their use of reference materials and their skills of reading information and applying it in written form and to develop a better understanding of and appreciation for the importance of women in our world. The skills included note taking, paragraphing, revising, and editing." This assignment was scored a "4" for coherence or match between learning goals and task since there is exact alignment between the teacher's stated goals and what the task asks students to do. The task fully supports the instructional goals.

Match Between Learning Goals and Grading
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	160502 500	This writing task was scored a 1 for match between learning goals and stated grading. The teacher's stated goals were "sentences, paragraphs, and writing creatively." The assignment required students to choose a character from a story, become that character, and describe their life as that character. The teacher's stated grading criteria upon which the students' work would be judged was the following list of questions: Did they choose a character? Did they take on the role? Did they follow instructions? The teacher added that creativity was not included in the criteria. The task was given this score because there is no alignment between the teacher's stated learning goals and the teacher's stated grading criteria. The teacher's goals were all centered around writing: sentences, paragraphs, and writing creatively. The criteria were all centered around whether the student was able to choose and become the character. There is no alignment between the two.
2	160501 500	In contrast to the previous scored assignment, there is a little alignment between the teacher's stated learning goals and the stated grading criteria in this task. Therefore it was scored a 2. The teacher's goals were "Brainstorming, drafting, learning to write report of information, and putting meaning to text." The teacher assessed students on "content and written expression." Both the goals and the grading criteria are vague, making it more difficult to determine the degree of match between the two. In the goals, the teacher wanted students to learn how to brainstorm, draft, and write an informational report. The teacher does not specify the actual skills involved such as finding resources, taking notes, judging what information is important to the report, writing paragraphs using one's own words, etc. The stated grading criteria, "content and written expression," are vague. Does content mean how much the student writes about the topic? Does written expression mean clarity? Does it mean writing in complete sentences? Does it include varying the beginning of sentences?

Score Point	Assignment ID	Assignment Description and Comments
3	160501 300	In this reading comprehension assignment, third-grade students were asked to write a retelling of the story *The Legend of the Persian Carpet* by Tomie de Paola. The teacher's goals were "to acquire comprehension skills and develop in writing skills as well. Most important is how much information they comprehend." The teacher used the following retelling rubric (Keene, Goudvis, & Schwartz, 1995): Random response; may be related to story (text); may give title Retelling reveals beginning awareness of event sequence Uses story elements (character, setting, conflict, sequence of events, resolution) and/or genre structure to organize a relatively accurate retelling (beginning, middle, end) Story elements/genre structure clear in an accurate retelling; refers to interactions between story elements (how problem affects character, how setting changes problem, etc.) Uses all story elements/genre structure and inferences to capture key themes in piece; points out interrelationships between elements; talks about how the overall meaning is influenced The rubric includes comprehension. It does not mention writing skills. The teacher's goals include both comprehension and writing skills, although the writing skills are not specified. This task was scored a 3 because there is only some or partial alignment between the teacher's stated learning goals and the stated criteria for grading. In addition, the rubric is not written in third-grade student language and therefore does not serve to communicate the teacher's expectations or goals clearly to the students.
4	160501 100	This task received a score of "4" for this dimension because there is exact alignment between the teacher's stated learning goals for students on this third-grade homework assignment and the teacher's stated grading criteria. The teacher's goals were "to have date, greeting, body and closing in proper places as well as writing sentences using capitals and punctuation accurately." According to the teacher's criteria, the students who received a high score used proper letter form, including a proper greeting and closing and indented paragraphs. Punctuation was mostly correct. Students who received an average or middle score used mostly proper form with some mistakes or they may not have indented but show the beginning development of paragraphs. They may or may not have used proper punctuation, and/or they closed the letter with "by" instead of "from, sincerely, etc." There is exact alignment between the stated goals and criteria, and the criteria include appropriate dimensions given the goals.

Overall
Third Grade

Score Point	Assignment ID	Assignment Description and Comments
1	N/A	N/A
2	010301 500	In this third-grade writing assignment, students were asked to write three sentences telling how the bear is going to get dinner. Students wrote a rough draft that was read and corrected by three peers before writing the final draft. The teacher's goals were: Writing a paragraph beginning with a rough draft and proofing by peers. Then writing a final draft for a rubric grade. The teacher used the school-developed third-grade writing rubric: 4 A Excellent—The writing is fluent and articulate. 3 B Outstanding—Good fluency and articulation. Few spelling errors. 2 C Satisfactory—Completion, writer exhibits some articulation and fluency. 1 D Not Satisfactory—Incomplete, difficult to understand. 0 F Unable to accomplish the task. This task was scored a 2 for overall quality because it is limited in terms of appropriateness of goal, application of goal in task, and scoring criteria. The teacher's purpose in giving this assignment is paragraph writing. This is an appropriate third-grade goal, but the application of the goal in the task is limited. This kind of writing assignment does not tap into the natural reasons for writing and does not tend to inspire excellent writing. The three sentences strung together to fulfill the task requirement do not hold together as a paragraph. In addition, the rubric is general and vague and not much help in giving feedback to students.
3	N/A	N/A
4	990502 600	In this third-grade challenging major assignment, students were asked to select a famous person to research and take notes on while reading in the following seven areas: birth and death, childhood, education, older years, why famous, other interesting information. The teacher's stated criterion for a high score was: Each of the seven sections contains accurate information and has several details above and beyond the minimum. The student uses correct sentence and paragraph structure In the original notes, the student did an excellent job summarizing the important information about the famous person. The criteria for a middle paper were: Most of the information is accurate. It may include minor errors. There is some information in each of the seven sections. There are at least three details for each section. It is written in complete sentences. The student at least attempts paragraph structure. This task was scored a 4 for overall quality because the teacher described clearly, specifically, and explicitly what is expected, the goals were appropriate, there was exact alignment between the goals, the task and the grading criteria, and the cognitive demands are appropriately challenging.

Cognitive Demands
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	100803 300	This sixth-grade reading comprehension assignment was scored a 1 for cognitive demand for seventh grade. Students were required to fill in blanks on a "basic recall worksheet" mostly created by the teacher. The questions required that students recount only basic facts about the story. The following are a few examples of the type of question asked in this assignment: (1) What was Jamie's first decision as treasurer? (p. 33); (2) How many people visited the museum on an ordinary Wednesday? (p. 36); (3) Why did Claudia and Jamie check their bags and instrument cases into the checking room? (p. 37); and (4) Find 3 words to describe Jamie's personality (p. 34, p. 34, p. 38). Students were generally given only one line on which to write their responses. Additionally, the teacher provided page numbers for the students to look up the answers. This further limited the level of challenge of this task since students were not expected to locate the answers from the text on their own or to recall the answers after hearing the chapter read aloud by the teacher. There were a total of 10 questions on the worksheet. The teacher said that two of the questions were discussed and answered as a group orally before students independently wrote down the answers. Very little thinking was required of students in this task.
2	011207 300	In contrast to the previous scored assignment, this reading comprehension assignment required students to write relatively more complete and detailed answers. Most of the questions students were asked only required recall basic facts; however, students were not given the page numbers of the answers. The following are examples of the type of question asked: (1) Why didn't the gods want humans to have fire?; (2) How did Prometheus plan to steal fire?; and (3) What happened to Prometheus after he stole the fire? Only one question required students to use moderately complex thinking and give a justification for their response: "Which theory do you think best explains the formation of the moon? Why?"
3	081403 300	In this homework assignment, students were asked to "reveal" a character from a book by creating a journal that would have been kept by that character. This assignment was scored a 3 for level of cognitive demand since it required that students infer a character's daily thoughts and feelings based on the character's experience in the novel. Rather than merely summarizing facts and events, students were required to apply and use the facts from the story in new ways (i.e., to create a new text), by determining and writing from the perspective of a single character.
4	160601 600	This challenging major assignment required students to create a newspaper set during WWII. This assignment was scored a 4 since it required that students bring together and analyze different types of information as demonstrated in their "cause and effect" essays and "evaluation." In addition to engaging with substantive content area knowledge (history), this assignment is also exceptional in that it required students to write in different styles (creatively, analytically, biographically, etc.), thus giving them experience writing in (and differentiating between) different genres.

Clarity of Teacher's Learning Goals
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	100805 500	This is an example of a 1 in terms of Clarity of Teacher's Goals. For this assignment, students were asked to evaluate a variety of children's books with mathematics as their central theme and choose a book that is suitable for children in K-4. The teacher's stated goals were as follows: *Evaluation of children's books, oral reading skills (using basic speech methods) and writing and illustrating a children's book.* It is nearly impossible to understand what the teacher had in mind for students to learn based on these objectives as they are stated. All except for perhaps the goals of [improving] oral reading skills are simply activities that students were required to perform. The objective of these activities is not known. There is no way to tell whether the teacher wished to improve students' math skills, their diction, their creativity in writing, their understanding of what makes a good children's book, etc.
2	081404 400	The following goals are an example of a 2 on the Clarity of Teacher's Goals scale. The assignment asked students to view a 30-minute episode of the movie "The Great Panda Adventure" and then write a summary of what they saw. The goals as stated by the teacher were as follows: *I expected students to be able to retell the sequence of events with a partner and then to write a brief summary of what happened.* These two goals are somewhat clear about what a student is supposed to gain—the skills of sequencing (or the ability to retell a story or the ability to work with a partner) and writing a summary. However, it is not clear what the final objective for student learning is—that is, what the student is supposed to be able to improve. Will this assignment help the student's writing ability, oral communication skills, analytical skills, memory?
3	081404 500	The following is an example of a 3 in terms of Clarity of Teacher's Goals. Students were to follow various steps in the writing process to write a report about mummification. The students worked from an outline that the class had generated together and followed a set of steps in producing the final draft: writing initial draft, revising, proofreading, peer reviewing, writing a second draft, teacher editing, publishing on the computer. The teacher wrote that the goals of this assignment were the following: *To improve the writing and thinking skills of drafting, revising, proofreading, editing, writing, and publishing.* These goals are clear and framed in terms of what students are to learn—the steps of the writing process. The separate goals or activities are very interrelated and work toward the general goal of improving a student's writing ability. However, this idea is not clearly communicated in the goals as they are stated, and the statement of the goals would have received a higher score if such elaboration had been included.
4	N/A	N/A

Grading Expectations
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	N/A	N/A
2	100805 300	The grading expectations for this seventh-grade reading comprehension assignment were scored a 2 because the teacher's scoring criteria were a list of undefined dimensions. The teacher gave a high score to papers with "complete sentences" and "well-thought-out answers," and papers that showed an understanding of poetry terms. Students who received a middle score wrote papers with some incomplete sentences and some answers that were not well thought out. In addition, these papers might also include some confusion of poetry terms. The scoring dimensions were not clearly defined for students. For example, it was not clear what the teacher expected would constitute a "well-thought-out" answer.
3	160601 500	In contrast to the previous scored assignment, the teacher on this challenging major assignment provided students with a much more detailed list of the elements that should be included in their evaluation of an American novel, and upon which they would be graded. For example, following is the list for the book report: 1. Introductory paragraph where you make a judgment based upon certain criteria 2. One paragraph where you develop your ideas, give examples, possible quotes, and your opinion 3. At least one more paragraph where you do the same as above 4. Closing paragraph where you close your essay 5. Interesting word choice and vocabulary 6. Language usage 7. Edited rough draft 8. Computer typed Following is the list for the "diamonte" (a summary of protagonist's character changes): 1. Interesting, descriptive word choices 2. Correct language choices 3. Neat and artistically decorated 4. Change in the protagonist 5. Rough draft included This task would have scored a 4 had the items been even more explicitly defined for students. For example, the teacher could have included some examples of "interesting, descriptive word choices." However, this list clearly communicated to students what they needed to include to obtain a good grade on the assignment. For this reason, this assignment was scored a 3 for teacher's grading expectations.

Score Point	Assignment ID	Assignment Description and Comments
4	160605 500	This challenging major assignment scored a 4 for teacher's grading expectations. Students were provided with an even more detailed list of what needed to be included in the assignment, and the criteria upon which their work would be assessed. Additionally, the teacher included a model of excellent work for the students. Following is the outline provided by the teacher for students to follow: I. Introduction A. Opening statement: Try to describe the book as best you can in one clear sentence. Include the title and author. (example: Pacific Crossing, by Gary Soto, is the story of one boy's experience during a summer spent in Japan. B. Explain the setting and give a little more detail on the story. C. Mention three important episodes from this story that you will talk about in the next three paragraphs. II. First episode A. Explain the beginning of the episode and give any background info needed to understand the episode. B. Summarize what happens and how the episode turns out. C. Tell what your main character learned from the events you describe. III. Second episode Follow instructions for paragraph II. IV. Third episode Follow instructions for paragraph II. V. Conclusion A. Say something about the book overall that connects this paragraph with the three that came before it. B. Explain what you think the important meaning of the book was, or what is valuable to learn from your book. C. Make a recommendation and give at least two reasons to support your recommendation. You can make a negative recommendation if you wish, just be sure you have two good reasons to support it. ("Dumb" or "boring" are not good reasons—you need to say why it is dumb or boring.)

Match Between Learning Goals and the Task
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	N/A	N/A
2	N/A	N/A
3	160605 200	This seventh- grade homework assignment asked students to read for 20 minutes and write a one-paragraph summary. The goals of the assignment are paragraph organization, comprehension development, and reading practice. This task was rated a 3 for coherence between Learning Goals and Task because there is alignment between the task and the goal. The task supports the learning goals. It is not a 4 because although the task supports the goal for the most part, it does not support the goal of comprehension development.
4	N/A	N/A

87

Match Between Learning Goals and Stated Grading
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	100804 400	This assignment asked students to write a business letter following a format, provided by the teacher, in which the students requested something from a company or individual. The teacher's stated learning goals are as follows: *Audience awareness, use of proper business language, business letter format.* The teacher stated the grading criteria as follows: *High (paper) would follow the format and show awareness of audience ("Waz up" is appropriate for a fun letter to Leo DiCaprio.). Middle (paper) didn't follow format so well, but was nonetheless thorough.* This is an example of a 1 because the grading criteria are not appropriate due to the lack of clarity. First, the criteria used for the high-quality papers are not consistently applied to the middle-quality papers. That is, different criteria are given for the different levels (i.e., awareness of audience vs. being thorough). Second, the letter that students were asked to write was a business letter, and it is inappropriate to consider in this category of letters a "fun letter."
2	081402 200	This assignment called for students to read a book set in the Middle Ages and then answer a prompt in which they analyze the differences between contemporary times and the Middle Ages. The learning goals and grading criterion as stated by the teacher were as follows: *Learning goal: The students are expected to be able to analyze the differences between their times and the times of their book. Grading criterion: Students who scored high were able to differentiate between the Middle Ages and contemporary times. They were able to elaborate on the differences and articulate them clearly.* The grading criterion here is not appropriate because it is so uninformative. The reader (or student) has no idea of the type of differentiations the teacher was looking for. An example from the teacher would have been useful here. Additionally, the teacher named a single criterion, and perhaps breaking down this general criterion into smaller criteria would have clarified the grading criteria.

Score Point	Assignment ID	Assignment Description and Comments
3	081404 400	The assignment for this example asked students to write a summary of events from a 30-minute sequence of the movie "The Great Panda Adventure." Before writing the summary, students were to retell the events of the story to each other. The learning goals as stated by the teacher were as follows: *I expected the students to be able to retell the sequence of events with a partner and then to write a brief summary of what happened.* The teacher's grading criteria were (1) students' ability to write a sequence of events, and (2) students' use of detail in the summary. This assignment is an example of a 3 in terms of the match between learning goals and grading criteria because the alignment between these is not complete. It is not clear from the teacher's statement of the grading criteria whether students are being held accountable for retelling the particular events from "The Great Panda Adventure" or merely writing any sequence of events.
4	N/A	N/A

Overall
Seventh Grade

Score Point	Assignment ID	Assignment Description and Comments
1	N/A	N/A
2	100805 400	In this seventh-grade task, students are provided with six pictures of a cat and a mouse. Students are to observe, discuss, and finally write a narrative based on these six pictures. The story could be from the point of view of the mouse or the cat. The goal of the assignment is to work on the narrative process. The teacher would like to get students to write in more detail about observable things and to write dialogue. Students were graded on the length of their stories, if they followed the sequence of the pictures, if they had a clear point of view and MUGS. This assignment was scored a 2 for overall quality for a couple of reasons. The challenge of the assignment does not seem developmentally appropriate seventh-grade students. While the goals themselves are clear, the grading criteria do not fit well with the goals of the lesson. Having a clear point of view seems to be the only aligned criterion for grading and goals.
3	160605 400	In this assignment, students were asked to read a book, at least 100 pages long, and to write a five-paragraph essay reviewing it. Students received an outline detailing what each paragraph of their review should contain. For example, in the introductory paragraph, the essay should contain: An opening statement in which students describe the book as best they can in one sentence, Explain the setting and give a little more detail on the story, Mention three important episodes from the story that you will talk about in the next three paragraphs. The goals for the task are writing from an outline, analyzing for theme/meaning, reading a whole book, and responsibility. Grading is based on adherence to the outline and MUGS. Students were also provided with a model of what is considered a high paper. This task received a score of 3 for overall quality. The goals are appropriate and are relevant to the task. The teacher provides students with a model of good work so they know what is expected of them. The grading criteria are clear and well defined. What keeps this task from being a 4 is the level of challenge. It is more focused on retelling the story than tapping into higher level thinking skills such as analyzing or making inferences.

Score Point	Assignment ID	Assignment Description and Comments
4	081402 400	Students were asked to analyze an article from an environmental magazine about an environmental problem. They were to write a letter to an individual or an organization that had some connection to the problem. In the letter, students were asked to describe the problem and pose possible solutions. The students were provided with guidelines describing specific criteria to include in the letter. Some kinds of the information students were to include were: the consequences of the environmental problem, the causes of the problem, and possible solutions. After students finished a draft of their letter, they gave the letter to another student who reviewed it using the same guidelines that were used to write the letter. Students then revised their letters. This task was scored a 4 because the assignment is challenging and rigorous, the goals are clear and aligned with the assignment, and the grading criteria are developed and include useful information for students about what is expected of them.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").